Evolutionary history and diversification mechanisms in the avian genus *Junco*: a multidisciplinary approach using phenotypic, ecological, phylogeographic and phylogenomic data

TESIS DOCTORAL
GUILLERMO TORGEIR FRIIS MONTOYA

Museo Natural de Ciencias Naturales
Universidad Autónoma de Madrid
Madrid, 2017

Thomas Nuttall, 1903

Museo Nacional de Ciencias Naturales
(Consejo Superior de Investigaciones Científicas)
Dpto. De Biodiversidad y Biología Evolutiva

Universidad Autónoma de Madrid
Dpto. De Biología

Evolutionary history and diversification mechanisms in the avian genus *Junco*:

a multidisciplinary approach using phenotypic, ecological, phylogeographic and phylogenomic data

Memoria presentada por el licenciado Guillermo Torgeir Friis Montoya para optar al grado de Doctor en Biología por la Universidad Autónoma de Madrid

DIRECTOR:
Dr. Borja Milá Valcárcel
Museo Nacional de Ciencias Naturales
Dpto. de Biodiversidad y Biología Evolutiva

TUTOR:
Dr. José Luis Bella Sombría
Universidad Autónoma de Madrid
Dpto. de Biología

TESIS DOCTORAL
Madrid, 2017

Cover:
Slate-colored junco illustration from "*A popular handbook of the birds of United States and Canada*"
by Thomas Nuttall and Montague Chamberlain, 1903

“Depuis sa naissance, dans les îles ioniennes, il y a près de trois mille ans, la pensée occidentale a été partagée entre deux attitudes en apparence opposes. Selon l’une de ces philosophies la réalité authentique et ultime de l'univers ne peut résider qu’en des formes parfaitement immuables, invariantes par essence. Selon l’autre, au contraire, c’est dans le movement et l’évolution que réside la seule réalité de l’univers.”

Jacques Monod, Le Hasard et la Nécessité (1970)

“We shall never fully understand nature (or ourselves), and certainly never respect it, until we dissociate the wild from the notion of usability.”

John Fowles, The Tree (1979).

Junco h. hyemalis

Junco h. aikeni

Junco h. oreganus

Junco h. mearnsi

Junco h. caniceps

Junco h. dorsalis

Junco p. palliatus

Junco p. fulvescens

Junco insularis

Junco p. alticola

Junco bairdi

Junco vulcani

CONTENTS

AGRADECIMIENTOS

Dicen que los agradecimientos, como las declaraciones de amor, deben ser totalmente convencionales; así pues, gracias a mi madre y a mi padre por haber hecho de mí la persona que soy, capaz de llevar a cabo esta tesis y de otros tantos partos, tanto o más esforzados. Gracias a mi hermano, que es sensible y valiente.

Gracias a mi director, Borja Milá, porque han sido su empeño y su integridad en el trabajo los que hicieron posible este proyecto, incluyendo mi beca predoctoral. Gracias por contar conmigo. Gracias por la compañía, el apoyo y los consejos durante todo el camino. Gracias a Pau Aleixandre y a Etienne Kornobis, que me dieron un recibimiento inmejorable y me ayudaron a dar los primeros pasos en el MiLab. Gracias a los que habéis venido después, Alvar Veiga y María Recuerda, espero que os sea tan grato como lo ha sido para mí. A mi tutor en la UAM, Pepe Bella, gracias por la atención y la buena disposición que siempre me has dedicado.

Gracias a mis amigos del MNCN, Paloma, Jorge, Miriam, Juanes, Cristina, Silvia, Paula, Ernesto, Samu, Lourdes, Martí, Merel, Miguel, Mar, Viole, Andrea, Buckley, Patri, Rodri, Mario, Iván. Todos habéis ayudado en algún momento, y no me refiero solo a la tesis. Gracias a Fergie por tu voluntarioso apoyo, y por nuestra incierta *astronomía de porvenires y recuerdos*. Gracias a Marta y Christoph, por organizar el Journal Club y por no dudar en aconsejarme y participar de nuestros proyectos. Gracias a Luisma Carrascal y Laura Barrios por prestaros a revisar mis análisis estadísticos y orientarme.

Fuera del Museo tengo también mucho que agradecer. A Modesto Luceño y el resto del grupo del área de Botánica de la Universidad Pablo de Olavide, el primer laboratorio con el que colaboré y en el que pude meterle mano a la molécula y a la Biología Evolutiva, muchas gracias. Gracias a Boris Delahaie, que por fortuna para mí andaba por Madrid cuando llegué y tanto ayudó con mis primeros análisis del color. Gracias a mis tutores y compañeros de las estancias y viajes, en especial a Amanda Zellmer y John McCormack, que me acogieron en su casa y que vienen colaborando y apoyando nuestro trabajo desde el primer día. También a su equipo, Whitney, James, Madhvi y los undergraduate students que tuve la suerte de conocer en mi etapa en OXY. Gracias a Bob Wayne y su grupo en UCLA, Jacqueline Robinson, Rachel Johnston, Sergio Nigenda, aprendí muchísimo con vosotros. Gracias a Benoit Nabholz y los estudiantes que conocí en el ISEM en Montpellier, Marjo, Clementine,

Yoann Anselmet y Yoann Ancieaux, Manon, fueron dos meses inolvidables. Gracias a Roy Hancliff, que muy amablemente me ha permitido usar sus magníficas fotos para ilustrar esta tesis

Tengo la suerte también de contar con varios colaboradores que no pueden faltar en estos agradecimientos. Además de Pau, es un honor firmar con Ricardo Rodríguez-Estrella y Adolfo G. Navarro-Sigüenza los resultados del Capítulo I, ya publicados; con Guillermo Fandos y Brant Faircloth, además de John y Amanda, todos colaboradores en el Capítulo III, que se publicará próximamente; y con Ángeles de Cara, colaboradora en el Capítulo IV, también pendiente de una pronta publicación. De igual forma, han sido de una inestimable ayuda en el desarrollo del proyecto Ellen Ketterson, Jonathan Atwell y Adam Fudickar, además de encantadoras personas. Hay mucha más gente, muchos de los cuales no he conocido, que vienen participando en el trabajo de campo necesario para este proyecto desde hace más de veinte años. A todos ellos, muchísimas gracias.

Quiero dar las gracias también a los miembros de mi tribunal de tesis por acceder sin la menor de las vacilaciones a formar parte del mismo: Maite Aguado, Christophe Thébaud, Íñigo Martínez-Solano y Cristina Grande. Mención especial a Carles Vilá, que ya estuvo en mi tribunal del TFM y que junto con Marta Barluenga y John McCormack, a quienes extiendo estos agradecimientos reiterados, se prestaron a formar parte del seguimiento anual de mi tesis en el 'Thesis Advisory Committee'.

Tengo que dar las gracias, finalmente, a los amigos que he hecho a lo largo de mi vida y que hoy unen fuerzas bajo diversas siglas: a los 'FlyMegaguay', a los 'MOMOS & TITIS', a la gran coalición 'MEGAFLY! deMOMOS, TITIS&RARRAS'. De verdad, gracias, sois de lo más preciado que tengo.

Por torpeza o por soberbia, por ingenuidad o por audacia me disculparán ustedes, ¿pueden creer que estuve tentado de despachar este asunto con un sumarísimo 'Gracias a todos'?

Evolutionary history and diversification mechanisms in the avian genus *Junco*: a multidisciplinary approach using phenotypic, ecological, phylogeographic and phylogenomic data

Doctoral candidate: Guillermo Torgeir Friis Montoya

Abstract

Recently diversified systems are optimal research subjects to study the relative roles of genetic drift, gene flow and selection in shaping patterns of diversity and promoting the formation of evolutionary lineages and species. In the early stages of the speciation process, causal correlations between patterns of phenotypic divergence and adaptive genetic variability with specific selective factors are still recent and detectable. In turn, reconstructing the evolutionary relationships of closely related lineages is a challenging task because of incomplete lineage sorting at many loci and potential gene flow among incipient lineages. In this thesis, I develop an experimental design that combines phylogenetic and phylogeographic analyses with phenotypic, ecological and genomic data in order to reconstruct the evolutionary history and study mechanisms of early lineage divergence in the songbird genus *Junco* (Aves: Emberizidae) from Central and North America. Previous analyses based on mtDNA markers revealed a lack of genetic diversity and strong signatures of recent population expansion in the phenotypically differentiated and geographically structured boreal forms of dark-eyed junco, suggesting a process of rapid diversification during a northward recolonization of the North American continent from Mexico after the last glacial maximum (LGM) ca. 18,000 years ago. Phylogenetic molecular dating confirmed the postglacial origin of the northern juncos, and phylogenomic analyses based on genotyping-by-sequencing (GBS) single nucleotide polymorphisms (SNPs) recovered a pattern of reciprocal monophyly among the rapidly diversified young lineages, contrasting with at least four isolated lineages in the south showing relative phenotypic stasis. Whole-genome analyses also supported the recent origin of the boreal forms of *Junco*, and recovered signals of demographic expansions and limited gene flow during lineage diversification, suggesting a scenario of rapid divergence in allopatry. Using linear regression and multivariate analyses, I also found signs of an

association between adaptive genomic variability and variation in both ecological and sexually selected traits, evidencing the roles of natural and sexual selection in jointly driving phenotypic diversification and lineage divergence in the recently diversified dark-eyed junco complex. In addition, genome surveys based on SNPs revealed genomic landscapes of divergence with no obvious regions of high differentiation, with significant outliers found to be scattered across the genome. These patterns are consistent with a diversification process driven by multiple selective factors acting on many independent genome-wide loci. Overall, the analyses and results reported in this dissertation reveal the *Junco* system as one of the fastest radiations known in vertebrates, driven by the combined effects of historical processes such as demographic expansions and drift in isolation, and also selective factors, including natural and sexual selection.

Key words: speciation, recent diversification, phenotypic diversification, postglacial expansion, avian radiation, natural selection, sexual selection, neutral divergence, gene flow

FOREWORD

The great diversity of life forms in our planet has challenged our understanding of the natural world and indeed the origin of our own species for centuries. Charles Darwin's decisive discovery of evolution through natural selection and the later development of the modern evolutionary synthesis established the theoretical basis of the mechanisms underlying the existence of discontinuous biological units. As a result, they provided an unequivocal explanation of the origin of biodiversity and of our relationship with the reality of the physical universe, shaping modern thought like no other scientific theory in history, in words of Jacques Monod.

The fundamental aim of evolutionary biology is thus to explain biological variation over time and space, and from the population to the species level. Two major tasks can be defined in order to study how this variability arises: (i) inferring the potential evolutionary relationships among the groups under study, so phenotypic differences and geographic patterns of variation can be analyzed in a historic framework; and (ii) inferring how selection, drift and gene flow shape variability within and among the groups under study, leading to lineage divergence. Since their appearance, molecular techniques have widely determined the extent to which these tasks were achievable. From population genetics to phylogeography and phylogenetics, three decades of molecular analysis have allowed to study the speciation process at different temporal and spatial scales, and to document the entire speciation continuum from early population divergence to full reproductive isolation. The arrival of high-throughput sequencing techniques has improved the resolution of the analyses and now phylogenetic reconstruction of both recent radiations and ancient lineages are within reach. Furthermore, we are now able to study how different evolutionary processes shape phenotypic variation at the genomic level, recovering in so doing the link between the evolutionary force, the evolving trait and its genetic basis. We are also starting to understand the complex genomic architecture of divergence and speciation. All these advances have greatly contributed to our knowledge of the evolutionary history of the organisms and of the evolutionary processes that gave rise to their astounding diversity.

Aiming to contribute to this knowledge, the objective of this dissertation is to reconstruct the evolutionary history and to study the evolutionary mechanisms

involved in the diversification of the genus *Junco* (Aves: Emberizidae), a case of rapid phenotypic differentiation across North America that provides a unique opportunity to test the relative roles of neutral and selective mechanisms in driving lineage divergence and speciation.

PRESENTACIÓN

La gran diversidad de formas de vida de nuestro planeta ha desafiado nuestro entendimiento del mundo natural y del origen de nuestra propia especie durante siglos. El decisivo descubrimiento de Charles Darwin de la evolución mediante selección natural y el posterior desarrollo de la nueva síntesis establecieron las bases teóricas de los procesos que dan lugar a la existencia unidades biológicas discontinuas. Como resultado, proporcionaron una explicación inequívoca del origen de la biodiversidad y de nuestra relación con la realidad del universo físico, contribuyendo como ninguna otra teoría científica a la construcción del pensamiento moderno, en palabras de Jacques Monod.

Así pues, el propósito fundamental de la Biología Evolutiva es explicar la diversidad biológica, tanto en el tiempo como en el espacio, y desde la población a la especie. Dos tareas principales pueden definirse a la hora de estudiar el origen de esta diversidad: (i) inferir las relaciones evolutivas potenciales entre los grupos de organismos bajo estudio, de forma que las diferencias fenotípicas y los patrones geográficos de variación puedan ser analizados en un contexto histórico; y (ii) inferir cómo la selección, la deriva y el flujo genético moldean la variabilidad dentro y entre los grupos de estudio, conduciendo a la divergencia de linajes. Desde su aparición, las técnicas moleculares han determinado en gran medida hasta qué punto estas tareas eran factibles. De la genética de poblaciones a la filogeografía y la filogenética, tres décadas de análisis molecular han permitido estudiar el proceso de especiación a diferentes escalas temporales y espaciales, además de documentar el 'speciation continuum' desde la divergencia poblacional temprana al completo aislamiento reproductivo. La aparición de las técnicas de secuenciación masiva ha incrementado notablemente la resolución de los análisis moleculares y a día de hoy la reconstrucción filogenética tanto de radiaciones recientes como de diversificaciones de linajes antiguos es posible. Más aún, hoy somos capaces de estudiar cómo diferentes procesos evolutivos contribuyen a la variación a nivel genómico, desentrañando por tanto la conexión entre la fuerza evolutiva, el rasgo en evolución y su base genética. Estamos también empezando a comprender la compleja arquitectura genómica que acompaña a los procesos de divergencia y especiación. Todos estos avances han contribuido enormemente a nuestro

conocimiento de la historia evolutiva de los organismos y de los procesos evolutivos que han dado lugar a su asombrosa diversidad.

Con el fin de contribuir a este conocimiento, el objetivo de la presente disertación doctoral es reconstruir la historia evolutiva y estudiar los mecanismos implicados en la diversificación del género Junco (Aves: Emberizidae), un ejemplo de rápida diversificación fenotípica en Norte América que supone una oportunidad única para testar los papeles relativos de mecanismos neutrales y selectivos subyacentes a la divergencia entre linajes y la especiación.

INTRODUCTION

Junco h. oreganus, picture by Roy Hancliff, www.royhancliff.com

The field of speciation research

The field of speciation has experienced a great development and an outburst of research in the last three decades, coinciding with advances in molecular techniques and a renewed interest in the mechanisms involved in lineage separation (Coyne and Orr 2004; Nosil 2012). Since Charles Darwin's publication of 'On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life' (1859) provided a gradualist perspective of change and speciation due to natural selection, the question of how species originate has received uneven attention up to the present day. During this time, numerous models of speciation have been proposed, and the specific problem of 'what is a species?' has received a number of answers. For the main part of the 20th century, the interbreeding-population concept that Ernst Mayr developed (1942), stood for (1963; Mayr and Ashlock 1969) and reviewed (1995), along with other prominent proponents like Dobzhansky (1935, 1950), has predominated. Under the term of 'Biological Species Concept' (BSC), this definition based on the requisite of reproductive isolation between biological species generated great consensus until the 80s. However, alternative species concepts have arisen in the latter decades to respond to different 'species problems', either to emphasize the evolutionary aspects in the species definition (Evolutionary Species Concept, Simpson 1961; Wiley 1978), pinpoint the mechanisms preserving phenotypic cohesion among members of a species (Cohesive Species Concept, Templeton 1989) or include ecological considerations in explaining discontinuities in sympatric sets of organisms (Ecological Species Concept, Van Valen 1976), to name a few. From the perspective of the evolutionary history of the species, one of the most informative definitions has been the Phylogenetic Species Concept (PSC), and more specifically, that proposed by authors like Rosen (1979) or Queiroz and Donoghue (Donoghue 1985; Queiroz and Donoghue 1988). The rise of the PSC is clearly associated with the beginning of the molecular era and the development of the molecular systematics, when nucleotide data could be used to define monophyletic groups based on shared derived characters.

In recent years, the debate over the 'species problem' has been, to some extent, overcome. Some authors like Jody Hey in his book from 2001, as well as John

Maynard Smith in the foreword, and Brookfield (2002) in the corresponding review, advocated a pragmatic use of the different species concepts in concordance with the question of study, and placed the discussion about how to define a species outside the limits of scientific research, and more related to philosophical concerns. Coyne and Orr adhered to this opinion in their book 'Speciation' (2004). In 2007, Queiroz proposed a unified species concept based on a common element, namely, that all definitions describe species as 'separately evolving metapopulation lineages'.

Parallel to the debate of the species problem, different modes of speciation focusing on different features of the speciation process have been proposed. Since Mayr's works, the fundamental and most discussed aspect of the speciation modes has been geography. During the times of the modern synthesis and up to our days, the allopatric model of speciation received abundant theoretical support and concentrated great consensus (Dobzhansky 1937; Mayr 1942, 1963). There is little controversy about the potential for populations in geographic isolation to differentiate by the action of *any* evolutionary force driving the development of reproductive isolation (Coyne and Orr 2004), including both neutral and selective processes. However, a debate remains open about how prevalent is the allopatric model with respect to other spatial modes of speciation. Many authors have defended that speciation is also possible in the presence of limited gene flow, a mode that in geographical terms was named 'parapatric speciation'. Populations connected by limited migration rates developing reproductive isolation due to selection have received verbal and mathematical support for both continuous populations distributed across selective gradients (Fisher 1930; Lande 1982; Barton and Hewitt 1985; Doebeli and Dieckmann 2002) and discrete populations experiencing low gene flow (Felsenstein 1981; Gavrilets 2000, 2004). Recent studies based on genomic analyses have reported conspicuous cases of parapatric speciation among closely related taxa (e.g. Nadeau et al. 2012; Andrew and Rieseberg 2013; Poelstra et al. 2014), reinforcing the hypothesis of speciation in the presence of restricted gene flow. In turn, the 'sympatric speciation' mode has generated less agreement (Coyne and Orr 2004), although theoretical support for speciation in the presence of extensive gene flow has been provided (Dieckman and Doebeli 1999; Gavrilets 2004; Kawecki 2004) and some examples in nature have

been reported (e.g. Feder et al. 1988; Sorenson et al. 2003; Barluenga et al. 2006), besides the well documented case of polyploid speciation in plants (Wood et al. 2009). The current predominant view is that sympatric speciation occurs in nature, especially through specific mechanisms like host shifts or instantaneous speciation, yet the frequency of sympatric speciation in nature is probably low (Coyne and Orr 2004; Bolnick and Fitzpatrick 2007).

In the last few decades, parallel to the surge in speciation research, speciation models have undergone a shift in their fundamental aspects, from the geographic factors of lineage splitting, to the specific mechanisms promoting reproductive isolation (Schluter 2000, 2001; Via 2001). Rather than focusing on the degree of spatial isolation, new models put emphasis on how gene flow and selection interact to produce phenotypic divergence and lineage diversification (Rundle and Nosil 2005; Nosil 2012). Geography remains an important component when analyzing the speciation process, yet redefined in function of its relationship with gene flow, that now is understood as a 'continuum' from complete isolation to panmixia (Butlin et al. 2008). The renewed interest in the role of selection in driving speciation has had relevant implications in theoretical and analytical approaches. It has paved the way for the 'ecological speciation' model, a mode of speciation by which new lineages form through cumulative, ecologically adaptive changes resulting in reproductive isolation (Mayr 1947; Schluter 2000; Rundle and Nosil 2005). This approach has gained support in the recent genomic era (Nosil 2012; Seehausen et al. 2014), yielding new explanatory models of patterns of genetic divergence in nature, as the 'isolation-by-adaptation' process (Nosil et al. 2008). Sexual selection has also regained popularity as a potential promoter of premating barriers (Lande 1981; West-Eberhard 1983; Panhuis et al. 2001; Kraaijeveld et al. 2011), a neglected aspect of evolution research during the times of the modern evolutionary synthesis. With the renewed emphasis on selection-based approaches to speciation research, adaptive radiations have also recovered popularity as model systems (e.g. Grant 1986; Schluter 1996; Wagner et al. 2012), and theory on the matter developed by Simpson (1953) has been thoroughly revisited and extended (Schluter 2000). The development of molecular techniques has also improved the analytical toolset to study selection's role in speciation. Phylogeography, originally based on neutrally-

evolving mtDNA sequence analysis, brought the formalization of conceptual links between intraspecific patterns of variation and evolutionary relationships among species (Zamudio et al. 2016), contributing to the understanding of microevolutionary processes leading to the formation of new species (Avise et al. 1987; Avise 2000). Molecular phylogenies and molecular dating also favored the study of organismal evolution in a historical framework (Avise 1994), enabled better informed comparative analyses of phenotypic evolution, enhanced the field of molecular systematics, and altogether helped to answer classical questions about the development of reproductive isolation and to interpret the relative roles of selection and drift in the speciation process (Coyne and Orr 2004).

Nowadays, the field of speciation research is undergoing a new transformation driven by the advance of genomics techniques. In the last ten years, high-throughput sequencing has profoundly enhanced our understanding of the speciation process across the whole speciation continuum. Thanks to these techniques, we are now able to investigate fundamental aspects of the genomic architecture of speciation and generate revised speciation models that include the number and localization of genes involved in reproductive isolation, the role of gene flow and introgression, and rate heterogeneity in divergence and recombination across the genome, among others (Stapley et al. 2010; Nosil 2012; Seehausen et al. 2014; Nosil et al. 2017; Riesch et al. 2017). Importantly, high-throughput sequencing has extended genomic-level speciation research to non-model organisms, enabling analyses in systems with different spatial settings, selection regimes and evolutionary histories, and allowing to test different hypotheses related to speciation. Many of these systems consist of young radiations, because of their suitability to detect evolutionary forces driving lineage formation and study speciation in action (e.g. Lerner et al. 2011; Keller et al. 2013). High-throughput sequencing markers have provided sufficient resolution to study speciation processes in these systems. As a result, the study of the early stages of speciation has become a central topic in the last few years of speciation research.

Speciation in birds

The class Aves is the most speciose of terrestrial vertebrates, and one of the best-known taxonomic groups. Birds are relatively non-elusive, and naturalists have been able to easily study them for centuries. Because the morphological and ecological diversity of birds is relatively well-known, the group has occupied a prominent position in the study of speciation since the dawn of evolutionary theory (Price 2008). The great diversity and conspicuousness of secondary sexual traits and courtships involved in mate choice behavior also provides excellent opportunities to explore the role of sexual selection in the speciation process. In addition, flight ability has allowed birds to colonize remote areas such as distant archipelagos and speciate in very specific conditions of geographic isolation and niche availability (Grant 1998; Lovette et al. 2002; Losos and Ricklefs 2009). Evolutionary biologists have long taken advantage of these systems to study speciation. Furthermore, the best-known avian adaptive radiations have taken place in isolated insular systems, and their study has greatly contributed to the development of models of speciation (Grant 2001). With the advent of genomic techniques, avian radiations have continued illuminating the processes leading to phenotypic divergence and lineage diversification.

The most celebrated and best studied adaptive radiation is that of Darwin's finches (Schluter 2000). The fifteen species of the genus *Geospiza* form a monophyletic group that evolved from an ancestor that colonized the Galápagos archipelago 2.3 million years ago ca. (Sato et al. 2001), and have evolved conspicuous diversity in beak shapes and sizes that correlate with the types of seed on which they feed (Grant 1986; Price 1987). In recent studies, genomic analyses have helped to identify the genetic basis of beak evolution in Darwin's finches (Lamichhaney et al. 2015; Chaves et al. 2016; Lamichhaney et al. 2016; Lawson and Petren 2017), revealing how selection acting on ecologically important fitness traits can shape the genomic landscape of divergence among lineages (Han et al. 2017).

There are also continental examples of adaptive radiations in terms of beak adaptations to food resources. The red crossbill complex (*Loxia curvirostra*) from North America may be subdivided into ten different categories depending on their call types (Groth 1993; Benkman et al. 2009; Irwin 2010) and also on the species of

conifers on which they forage (Benkman 1987, 1993, 1999). In a thorough study by Benkman and collaborators (2003), different types of crossbill were shown to reside near different optimal phenotypic peaks of fitness depending on the food resource the exploit, and how these adaptations can drive reproductive isolation among types in relatively short periods of time. Analyses based on genome-wide single nucleotide polymorphisms (SNPs) revealed a pattern of shallow genetic structure despite the absence of geographic barriers, congruent with a process of differentiation as a byproduct of adaptation to specific food resource (Parchman et al. 2016). Furthermore, a population of resident crossbills in the South Hills presented relatively high genetic divergence in a few loci that may be involved in a coevolutionary arms race, contributing to genome-wide divergence in periods as short as 6,000 years (Parchman et al. 2016).

Many other avian systems have contributed to our understanding of the speciation process. The Pleistocene radiation of the white-eyes (genus *Zosterops*), known as the 'great speciators', revealed a pattern of macroevolutionary diversification determined by lineage-specific life-history traits over global geographic regions, and a remarkable capacity to reach far-away islands, with over 40% of the species being single-island endemics (Moyle et al. 2009). The South American *Sporophila* seedeaters in turn, present up to eight sympatric, markedly differentiated types of plumage coloration originated in the last 40,000 years, with very low genetic variation (Campagna et al. 2015; Campagna et al. 2017) suggesting a role for sexual selection driving phenotypic diversification. Avian species like the great tit (*Parus major*) or the zebra finch (*Taeniopigya guttata*) have served as model species in neuroscience or behavior research (Griffith and Buchanan 2010; Warren et al. 2010), and groups like the birds of paradise (Paradisaeidae) have profoundly fascinated and challenged many evolutionary biologists in their research about the role of sexual selection in promoting evolutionary change (Barker et al. 2004; Scholes 2008). With the arrival of the genomic era, many avian non-model species have become excellent systems for speciation research.

High-throughput sequencing and the power of multidisciplinary approaches in speciation research

Since the first high-throughput sequencing platform was released (Roche 454, Margulies et al. 2005), the dramatic decrease in sequencing costs has permitted the access to genomic techniques for many laboratories and research groups. The platforms by Illumina Inc. have been especially relevant in providing cheaper sequencing services and they are the most widely used high-throughput sequencers nowadays (Bleidorn 2017). While this technology made feasible the sequencing of full genomes, for many studies in speciation research, sequencing of reduced fractions of the genome for a large number of individuals may be efficient and more affordable (Andrews et al. 2016; Bleidorn 2017). Restriction-site associated DNA sequencing (RAD-seq) methods have been particularly useful for genome-wide SNP discovery, and widely applied in the study of population structure and the reconstruction of the evolutionary history of young taxa, because of their high potential to detect phylogenetic signal and avoid problems like incomplete lineage sorting or lack of resolution at recent evolutionary scales (Cariou et al. 2013). The first RADseq method was developed by Baird et al.(2008) in order to sample genome-wide SNP variation in RAD sequences using short-read high-throughput technology like Illumina. Later, Emerson et al. (2010) used RAD sequencing to identify the previously difficult to resolve phylogeographic relationships among postglacial populations of the pitcher plant mosquito. Since Emerson et al.'s publication several studies using RADseq and similar genotyping-by-sequencing (GBS, Elshire et al. 2011) techniques for population genomics (e.g. Hohenlohe et al. 2010) and shallow phylogenetic reconstruction (e.g. Jones et al. 2013; Wagner et al. 2013) have appeared. Reduced-representation genome-sequencing techniques have thereby fostered considerable progress in our analytical capacity in population genetics, phylogeography and phylogenetics (McCormack et al. 2013; Edwards et al. 2016; Zamudio et al. 2016; Bleidorn 2017), but also in genome scale analyses for detecting the genetic basis of adaptive divergence (Stapley et al. 2010; Rice et al. 2011; Andrews et al. 2016). For instance, Hohenlohe et al. (2010) were able to identify candidate loci for stickleback phenotypic evolution using RADseq, and Ruegg et al. (2014) found signs of increased differentiation in genes linked to migration behavior of Swainson's thrush.

Despite the power of RADseq methods, as modern Illumina platforms (e.g. HiSeq or the recently released NovaSeq) keep reducing the costs of sequencing, the number of full genome based studies in speciation research has increased. Whole-genome sequencing yields better resolution than RADseq methods and thus provides a better basis for the detection of candidate genes, the study of the role of genomic architecture and structural variation involved in lineage formation and divergence, or testing for signatures of introgression, to name a few applications (Seehausen et al. 2014). Thanks to sequencing strategies like whole-genome shotgun, resequencing is becoming a progressively viable option for non-model organisms, and as a result, a number of highly relevant studies in the field of speciation research based on full genome analyses have appeared in the last years (e.g. Martin et al. 2013; Poelstra et al. 2014; Soria-Carrasco et al. 2014).

In the last decade, high-throughput sequencing has profoundly enlarged our knowledge of the mechanisms of speciation by providing new analytical tools. Besides improving phylogenetic analyses by making available large numbers of markers, tools like landscape genomic surveys based on both RADseq (e.g. Hohenlohe et al. 2010; Keller et al. 2013) and whole-genome data (e.g. Jones et al. 2012; Poelstra et al. 2014) have greatly contributed to our understanding of how evolutionary forces shape genome-wide divergence, how many loci may be involved in lineage diversification, and how they are distributed across the genome (Feder et al. 2012; Burri et al. 2015; Riesch et al. 2017). Genome scans have also been useful for identifying highly differentiated loci, potentially under divergent selection, as candidate genes involved in reproductive isolation (Faria et al. 2014; Rellstab et al. 2015). Genomic tests to explicitly measure and explore the role of gene flow and hybridization in speciation have also been developed, such as the ABBA-BABA test (Green et al. 2010; Durand et al. 2011) that was originally used to explore hybridization between humans and Neanderthals, but successfully applied to different biological systems (e.g. Martin et al. 2013; Rheindt et al. 2013). New Bayesian, coalescent-based methods like G-PhoCS (Generalized Phylogenetic Coalescent Sampler, Gronau et al. 2011) can estimate the rates of gene flow among branches of a given phylogeny while inferring ancestral population sizes using

genome-wide neutral loci (e.g. Freedman et al. 2014; Campagna et al. 2015). Other methods for estimating ancestral population sizes based on genomic markers are ∂a∂I (Gutenkunst et al. 2009), or MSMC (multiple sequentially Markovian coalescent, Schiffels and Durbin 2014), also developed for humans in origin, but applied to different organisms (e.g. Freedman et al. 2014; Bosse et al. 2015; Eaton et al. 2015).

Recently developed approaches in the field of molecular ecology have taken advantage of high-throughput sequencing to study how ecologically based selection and local adaptation drive lineage diversification. These methods look for correlations between allele frequencies and phenotypic measurements, like in genome-wide association studies (GWAS, reviewed in Korte and Farlow 2013) and quantitative trait locus (QTL) mapping (reviewed in Stinchcombe and Hoekstra 2008). They can also estimate statistical association patterns between genetic variance and variation in environmental parameters, in order to identify ecological factors driving differentiation due to local adaptation, a set of methods known as genetic-environment association methods (GEA, Frichot et al. 2015; Rellstab et al. 2015; Forester et al. 2016). Importantly, GEA approaches allow the inclusion of covariables in order to control by a number of other effects, like population history or geography, making them powerful tools in the study of local adaptation and speciation. Programs like Bayenv (Coop et al. 2010) or the more recent BayeScEnv (Villemereuil and Gaggiotti 2015) are also intended to find signs of selection driven by specific climatic variables, integrating both genomic and ecological data in a Bayesian analysis framework.

A myriad of other genomic tests along with their variants with applications in evolutionary biology and speciation research have been developed, giving birth to the discipline known as 'speciation genomics'. Analyses integrating genomic markers and both ecological and phenotypic data have been particularly useful to study the role of adaptive divergence in driving lineage diversification, especially at early stages of the process (Stinchcombe and Hoekstra 2008; Faria et al. 2014; Seehausen et al. 2014; Grant and Grant 2017), providing great insight in the

mechanisms underlying speciation (e.g. Wagner et al. 2012; Linnen et al. 2013; Szulkin et al. 2016).

Study System

This thesis takes advantage of the newly developed high-throughput sequencing techniques and of an extensive sample collection of genetic, phenotypic and ecological data to study the early stages of the speciation process in the songbird genus *Junco*. The wide range of patterns of geographic variation found in the juncos has raised numerous studies and debates about their evolutionary history long before molecular techniques were developed and applied to the reconstruction of phylogenies. The most remarkable one is the monograph "Speciation in the avian genus Junco", by Alden Miller (1941). In this volume, Miller identified and described a total of 21 different "forms" or "types", and grouped them in ten species based on their phenotypic similarities, geographical distribution and potential evolutionary relationships. However, because several of those forms were found to interbreed freely in areas of parapatric contact, the American Ornithologists' Union currently groups those 21 forms into just five species: the divergent *Junco vulcani* in the highlands of Costa Rica; *Junco bairdi* from the southern tip of the Baja California Peninsula (recognized recently based on the work developed in this thesis); the island junco *Junco insularis* on Guadalupe Island; the yellow-eyed junco *Junco phaeonotus* in the highlands of Mexico and Guatemala; and the dark-eyed junco *Junco hyemalis*, which inhabits conifer and broadleaf forests across temperate and boreal North America (Sullivan 1999; Nolan et al. 2002). The yellow-eyed junco is divided into five moderately differentiated subspecific taxa: *J. p. palliatus* of northern Mexico, *J. p. phaeonotus* of central Mexico, *J. p. fulvescens* of Chiapas in southern Mexico, *J. p. alticola* of Guatemala, and *J. p. bairdi* of the southern tip of the Baja California Peninsula. In contrast, the dark-eyed junco presents a striking diversity of plumage patterns and colors, and is composed of at least 14 distinct and largely allopatric morphotypes classified into four major groups: the slate-colored junco [*hyemalis* group] in eastern and boreal North America, composed of three subspecific taxa, *J. h. hyemalis*, *J. h. carolinensis* and *J. h. cismontanus*); the white-winged junco *J. h. aikeni* in the Black Hills of South Dakota; the Oregon junco

[*oreganus* group] across the West, composed of several distinct forms from Alaska to northern Baja California, including *oreganus*, *shufeldti*, *montanus*, *thurberi*, *pinosus*, *pontilis*, *townsendi* and the pink-sided junco *mearns*i in the northern Rocky Mountains; and the gray-headed junco [*caniceps* group] in the Rocky Mountains and southwestern USA, composed of *J. h. caniceps* and *J. h. dorsalis* (Fig. 1, Table 1).

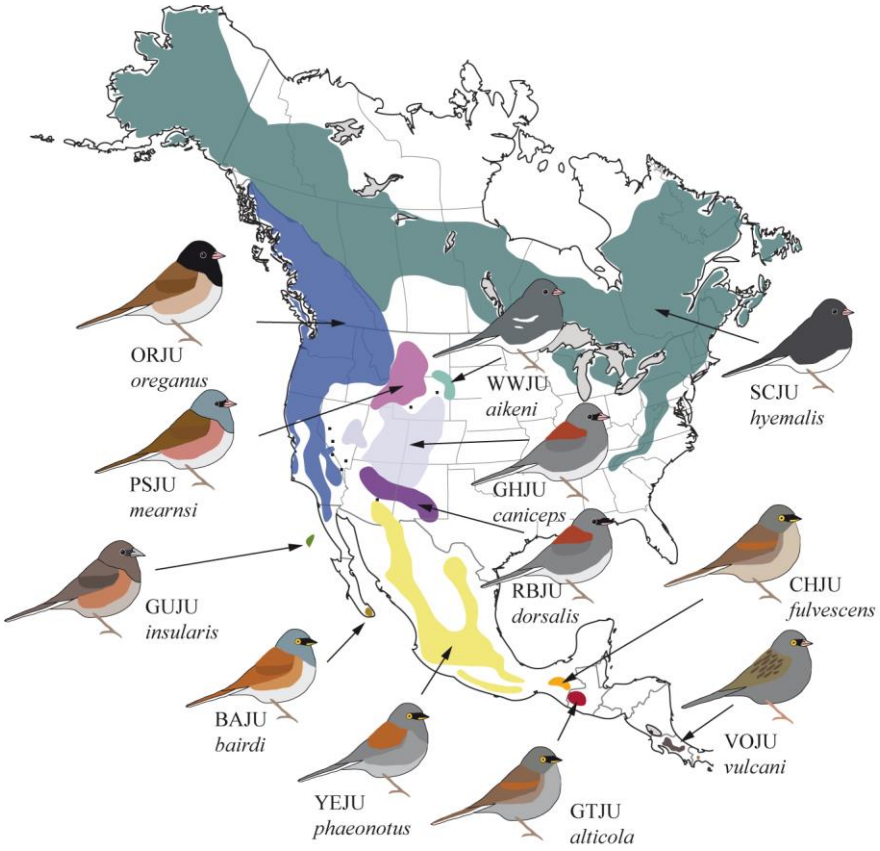


Figure 1. Distribution map of the different junco forms. Colored areas correspond to the breeding ranges of the major forms (see Table 1 for a detailed nomenclature). Dots represent isolated localities with hybrid/intermediate individuals.

In a previous study, Milá et al. (2007) reported for the first time spatial patterns of genetic diversity in several forms of Central and North American juncos, along with a haplotype network based on neutral molecular markers and an analysis of population structure based on amplified fragment length polymorphism (AFLP) loci (Fig.2). The study revealed a lack of genetic diversity and strong signatures of recent population expansion in the northern forms. The population structure analysis also revealed a gradual pattern of differentiation between yellow-eyed and dark-eyed

populations, with assignment probabilities changing with latitude. As pointed out by Milá et al. (2007), this diversity pattern suggests a complex scenario of rapid diversification of the northern forms, seemingly occurred during a northward expansion after last glacial maximum ca 18,000 years ago. This contrasted notably with patterns of phenotypic and genetic differentiation in the forms from Guatemala (*alticola*), and Guadalupe Island (*insularis*), which despite their more discrete divergence in plumage color, turned out to be highly isolated lineages an order of magnitude older, as shown in a later study by Aleixandre et al. (2013), also based on mtDNA markers.

This striking pattern of geographic variation of the juncos across a complex array of southern divergent lineages and young, recently diversified forms offer a unique case of 'speciation continuum', from the local population to the fully isolated species, exemplifying the different stages of the lineage divergence process in different spatial settings that encompass everything from completely allopatric species to parapatric, interbreeding forms (Miller 1941; Milá et al. 2016). Only in the dark-eyed junco species, there are several degrees of differentiation and ongoing gene flow among forms. We can find highly dissimilar allopatric forms like the slate-colored (*hyemalis*) and red-backed juncos (*dorsalis*); there are geographically isolated forms in continental sky islands like those from northern Baja California (townsendi and pontilis), restricted to high altitudinal patches of forest; we also find forms presenting different degrees of hybridization with neighboring forms, such as Oregon juncos, which mix with the gray-headed juncos at narrow and discontinuous contact zones in California and Nevada (thurberi and caniceps), yet show extensive interbreeding with the slate-colored juncos across broad regions in western Canada (montanus and oreganus, hyemalis and cismontanus) (Milá et al. 2016). The complexity of the Junco system provides therefore an excellent opportunity as a model for the study of speciation processes

Figure 2. Geographic scope and genetic analyses from the study by Milá et al. (2007). (A) Sampling localities across different junco forms. The eight different taxa included in the genetic analyses are shown (see Table 1 for a detailed nomenclature). Site 23 corresponds to a wintering locality of the slate-colored junco in Alabama. (B) Posterior assignment probabilities of 139 individuals in different Junco populations to K values of 2, 3 and 4 using 75 AFLP loci in the program STRUCTURE. Each color in each pie diagram represents the percent posterior probability of assignment to a given cluster, averaged across all individuals in that population. (C) Minimum-spanning network of absolute distances between mtDNA control region haplotypes found in 264 individual juncos. Each circle represents a haplotype, with size proportional to the haplotype's overall frequency. Letters designate haplotypes and numbers correspond to the sampling localities where the haplotype was detected. Network branches represent a single nucleotide change and hatch marks along branches represent additional changes. Modified from Milá et al. 2007.

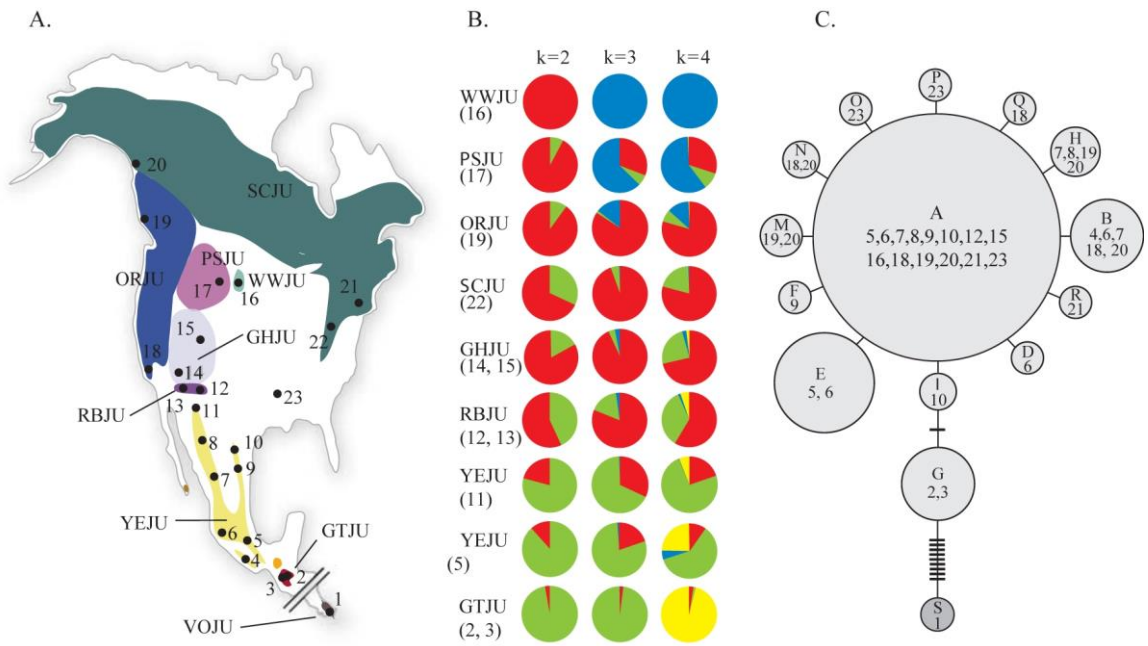


Table 1. Taxonomy of junco forms based on Miller (1941). The five taxa reported under "Species" are those currently recognized by the American Ornithologists' Union (Chesser et al. 2017).

Country	**Species**	**Forms**	**Common name**	**Code**	
USA and Canada	*J. hyemalis*	*hyemalis* *cismontanus* *carolinensis*	Slate-colored junco	SCJU	Dark-eyed junco (DEJU)
		aikeni	White-winged junco	WWJU	
		caniceps	Gray-headed junco	GHJU	
		dorsalis	Red-backed junco	RBJU	
		mearnsi	Pink-sided junco	PSJU	
		oreganus *shufeldti* *montanus* *pinosus* *thurberi*	Oregon junco	ORJU	
Mexico		*pontilis* *townsendi*			
	J. insularis	*insularis*	Guadalupe junco	GUJU	
	J. bairdi	*bairdi*	Baird's junco	BAJU	Yellow-eyed junco (YEJU)
	J. phaeonotus	*palliatus*	Yellow-eyed junco	YEJU	
		phaeonotus			
		fulvescens		CHJU	
Guatemala		*alticola*		GTJU	
Costa Rica	*J. vulcani*	*vulcani*		VOJU	

GENERAL OBJECTIVES

The purpose of this thesis is to reconstruct the evolutionary history of the genus *Junco* (aves: Emberizidae) and to study the causal evolutionary forces and mechanisms of phenotypic divergence and lineage diversification in this non-model biological system, with special emphasis on the early stages of the speciation process, and by implementing a multidisciplinary approach based on phenotypic, ecological and genomic data. My first goal is to use single-nucleotide polymorphism (SNP) markers and phylogenomic techniques to reconstruct the phylogenetic relationships among junco forms, and then integrate a multidisciplinary approach using phenotypic, ecological and genomic data to understand the evolutionary forces underlying geographically structured variability across the genus. In **Chapter I**, I implement the genotyping-by-sequencing (GBS) technique to generate thousands of SNPs for recovering the evolutionary history of the group, especially the relationships among *Junco hyemalis* forms, which diversified after the last glacial maximum (LGM) and cannot be resolved with mtDNA markers. In light of this new dataset, I also analyze the variability in morphology and plumage coloration to assess the role of adaptation and sexual selection in the radiation of the genus. In **Chapter II,** I explore the geographic variation of phenotypic traits (morphology and plumage coloration) to assess the role of sexual selection in driving differentiation of the various species and subspecies within the genus. In **Chapter III**, I apply genetic-environment association (GEA) analyses integrating genomic and ecological data to test the role of adaptation in driving population differentiation in the relatively homogeneous, widely distributed Oregon junco of western North America. In **Chapter IV,** I combine whole-genome and GBS data to reconstruct the demographic history of the *Junco hyemalis* forms to test the hypothesis of diversification process during the northward demographic expansion after the LGM. Specifically, I test the population-expansion in northern junco forms using G-PhoCS (generalized phylogenetic coalescent sampler), and explore how genomic patterns of variation relate to demographic events by means of genomic landscapes of variation.

GENERAL METHODS

A number of methods and analyses are recurrent through all the Chapters included in this dissertation. For the sake of clarity and brevity, I report them here, and subsequent chapters will refer to this general section when needed.

Population sampling

Juncos were sampled across their range using mist nets in order to obtain biological samples for DNA extraction. Each captured individual was aged, sexed, and marked with a numbered aluminum band. A blood sample was collected by venipuncture of the sub-brachial vein and stored in Queen's lysis buffer (Seutin 1991) or absolute ethanol at -80ºC in the laboratory. After processing, birds were released unharmed at the site of capture. All sampling activities were conducted in compliance with Animal Care and Use Program regulations at the University of California Los Angeles, and with state and federal scientific collecting permits in the USA and Mexico. Tissue samples from two individuals of *Junco vulcani* were kindly loaned by the Louisiana Museum of Natural History at Louisiana State University (catalog numbers: LSU16242, LSU16243). A high-quality tissue sample for whole-genome sequencing was provided by the Moore Laboratory of Zoology, Occidental College (specimen #69090). For a complete list, see Appendix I of this section. Genomic DNA was extracted from blood and tissue samples using a Qiagen DNeasy kit (Qiagen™, Valencia, CA) for downstream analyses.

Phenotypic data and analysis

We obtained morphometric and colorimetric data from 639 museum specimens representing all main junco forms, deposited at various natural history museums (see Appendix I). A wing ruler was used to measure unflattened wing length to the nearest 0.5 mm, and dial calipers of 0.1-mm precision were used to measure tail length, tarsus length, bill culmen, exposed bill culmen, and bill width and depth, following Milá *et al.* (2008). All measurements were taken by a single observer (Borja Milá).

To collect reflectance spectra we used a JAZ-EL200 spectrophotometer with a deuterium-tungsten light source via a bifurcate optical fiber probe (Ocean OpticsTM). The reflectance captor probe was mounted on a black rubber holder which excluded all external light and maintained the probe fixed at a distance of 3 mm from the feather surface at a 90° angle (e.g. Schmitz-Ornes 2006; Chui and Doucet 2009). The spectrum of each measurement ranged from 300 to 700 nm and consisted of three replicate measurements of three different readings per replicate, taken on each of six plumage patches: crown, nape, back, breast, flank and belly. Replicates were averaged before analysis. All reflectance data are expressed as the percentage of reflectance from a white standard (WS-1, Ocean OpticsTM). The white standard was measured after each specimen and the spectrophotometer was recalibrated regularly. All measurements were taken by myself.

We obtained colorimetric variables by applying the avian visual model by Stoddard and Prum (2008), based on Goldsmith's (1990) tetrahedral color space for spectral data. We used the R-package pavo (Maia et al. 2013) to calculate the relative quantum catch for each cone using the function vismodel. Specifically, we applied the visual system, sensitivity and ocular environmental transmission of the blue tit as available in the package, the 'forestshade' illuminant option and an ideal homogeneous illuminance for the background. We also applied the von Kries color correction transformation. We then obtained the spherical coordinates of tetrahedral color space describing the hue (Θ and φ) and the achieved chroma (r_a) using the function colspace. We included the normalized brilliance as a fourth variable, computed as described in Stoddard and Prum (2008). All the analyses were run in R Studio (R Studio Team 2015) version 1.0.136 with R (R Core Team 2015) version 3.2.2. Corresponding R code included in Annex I.

Whole genome sequencing and assembly

We used eight different full genome sequenced samples of different forms of junco for assembling a consensus draft genome to be used as a conspecific reference in the SNP calling process. Libraries for seven of the genomes were prepared with the Kapa Library Preparation Kit (Kapa Biosystems, Inc.) using TruSeq-style adapters

(Faircloth and Glenn 2012). They corresponded to the forms *bairdi* (n = 2, sampled in Sierra de la Laguna, Baja California, Mexico), *palliatus* (n = 2, sampled in Tescate, Chihuahua, Mexico), *caniceps* (n = 2, sampled in the Toiyabe Mountains, Nevada, USA) and *thurberi* (n = 1, for a second sample sequencing failed, sampled in Mount Laguna, California, USA). They were pooled after random shearing and individual barcoding and sequenced in a single lane of an Illumina HiSeq platform. The eighth genome, a *dorsalis* specimen (n = 1, Dude Mountain, Arizona, USA; skin stored in the Moore Laboratory of Zoology, Occidental College, specimen #69090) was sequenced at a higher coverage by means of two 101-bp paired-end shotgun libraries and two 101-bp mate-paired libraries with insert sizes of 8 Kb in length at Macrogen Inc. The TruSeq Nano DNA Kit (Illumina) was used for the preparation of the shotgun libraries, while the mate-paired libraries were prepared with Nextera Mate Pair Kit (Illumina). We used FASTQC (Andrews 2010) to evaluate the quality of the sequenced data, and quality filtering was carried out with NextClip (Leggett et al. 2013) in the case of the mate paired libraries. For the rest of them we used Trimmomatic (Bolger et al. 2014), applying a sliding window filtering approach with a size of 4pb and a phred quality score threshold of 25. We also set a minimum length of 50 bp, below which reads were filtered out after trimming. We used the software SOAPDENOVO2 (Luo et al. 2012) to perform the assembly. The average insert size for each library was estimated in a preliminary run, and we set a Kmer size of 27 and minimum edge coverage of 2. Gaps emerged during the scaffolding process were removed with the GapCloser tool from SOAPDENOVO2. Finally, we filtered out all the scaffolds shorter than 500 bp so the genome was functional as a mapping reference. The final assembled genome had 37,904 scaffolds with an N50 of 147,816 bases, a L50 of 1,951 scaffolds, 1.09 Gb length, 17.5 Mb of missing sites and an overall coverage of ~56X, as computed with VCFTOOLS version 0.1.13 (Danecek et al. 2011).

Genotyping-by-sequencing

We used genotyping-by-sequencing (Elshire *et al.* 2011) to sequence five plates, each one consisting of 95 individually barcoded DNA samples and a blank, for a total of 475 samples genotyped. Out of these, 21 corresponded to repeated individuals to

be used as controls in the SNP calling process, so our final dataset consisted of 454 individual genotypes, belonging to the following forms (with sample sizes in parentheses): *vulcani* (2), *bairdi* (8), *alticola* (8), *insularis* (7), *fulvescens* (3), *phaeonotus* (7), *palliatus* (9), *dorsalis* (50), *caniceps* (79), *townsendi* (16), *pontilis* (16), *thurberi* (35), *pinosus* (16), *shufeldti* (12), *montanus* (19), *oreganus* (16), *mearnsi* (17), *aikeni* (17), *carolinensis* (22), *hyemalis* (14) and distinct hybrid individuals (81) (See Appendix II). GBS libraries were prepared and sequenced at Cornell University's Institute for Genomic Diversity, using the restriction enzyme PstI for digestion (Baldassarre et al. 2014; Harvey and Brumfield 2015). Each plate was sequenced in a different lane of an Illumina HiSeq 2000 platform, resulting in 191.5 to 276 million good single-end reads of 101 bp in length per lane, and from 17.4 to 23.2 million individual tags (unique sequences from one or more barcoded reads, Glaubitz et al. 2014) (see Table 1 from this section for per plate details).

Table 2. Summary of reads and tags found in each GBS plate, corresponding to a single lane in an Illumina Hiseq sequencer.

Plate	Reads	Good Reads	Tags
Plate 1	255,680,820	191,478,001	28,414,236
Plate 2	293,426,436	268,302,771	22,491,338
Plate 3	275,051,299	252,224,543	20,426,202
Plate 4	301,301,828	275,974,867	23,225,951
Plate 5	246,061,872	227,802,346	17,399,836

Alignment and variant calling

We evaluated GBS read quality using FASTQC after sorting them by individual with FASTXTOOLKIT (Gordon and Hannon 2010), and performed the trimming and filtering treatment using PRINSEQ (Schmieder and Edwards 2011). All resulting reads were 69 bp long and had a mean genotyping phred quality score of at least 30, with no positions below 20. The reads were then mapped against the assembled junco genome (except for analyses in Chapter I) using the mem algorithm in the Burrows-Wheeler Aligner (BWA, Li and Durbin 2009). We used the Genome Analysis Toolkit (GATK, McKenna et al. 2010) version 3.6-0 to realign reads around

indels using the IndelRealigner tool and then we applied the HaplotypeCaller tool to call the individual genotypes. We finally used the GenotypeGVCFs tool to gather all the per-sample GVCFs files generated in the previous step and produce a set of joint-called SNP and indel (GATK Best Practices, DePristo et al. 2011; Auwera et al. 2013) in the variant call format (*vcf*). Because GBS data do not provide enough coverage for base quality score recalibration, we used VCFTOOLS to implement a 'hard filtering' process, customized for each of the downstream analyses (see Chapters).

References

Aleixandre, P., J. Hernández-Montoya, and B. Milá. 2013. Speciation on oceanic islands: rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: *Junco*). PLOS One 8(5):e63242.

Andrew, R. L. and L. H. Rieseberg. 2013. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution 67:2468-2482.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews. Genetics 17:81.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, and J. Thibault. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current protocols in bioinformatics:11.10. 11-11.10. 33.

Avise, J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489-522.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. Plos One 3:e3376.

Baldassarre, D. T., T. A. White, J. Karubian, and M. S. Webster. 2014. Genomic and Morphological Analysis of a Semipermeable Avian Hybrid Zone Suggests Asymmetrical Introgression of a Sexual Signal. Evolution 68:2644-2657.

Barker, F. K., A. Cibois, P. Schikler, J. Feinstein, and J. Cracraft. 2004. Phylogeny and diversification of the largest avian radiation. Proceedings of the National Academy of Sciences of the United States of America 101:11040-11045.

Barluenga, M., K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439:719.

Barton, N. H. and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annual Review of Ecology and Systematics 16:113-148.

Benkman, C. W. 1987. Crossbill foraging behavior, bill structure, and patterns of food profitability. The Wilson Bulletin:351-368.

Benkman, C. W. 1993. Adaptation to Single Resources and the Evolution of Crossbill (Loxia) Diversity. Ecological Monographs 63:305-325.

Benkman, C. W. 1999. The selection mosaic and diversifying coevolution between crossbills and lodgepole pine. the american naturalist 153:S75-S91.

Benkman, C. W. 2003. Divergent selection drives the adaptive radiation of crossbills. Evolution 57:1176-1181.

Benkman, C. W., J. W. Smith, P. C. Keenan, T. L. Parchman, and L. Santisteban. 2009. A new species of the red crossbill (Fringillidae: Loxia) from Idaho. The Condor 111:169-176.

Bleidorn, C. 2017. Phylogenomics. Springer International Publishing.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics:btu170.

Bolnick, D. I. and B. M. Fitzpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459-487.

Bosse, M., H.-J. Megens, O. Madsen, R. P. M. A. Crooijmans, O. A. Ryder, F. Austerlitz, M. A. M. Groenen, and M. A. R. de Cara. 2015. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome research 25:970-981.

Brookfield, J. 2002. Genes, Categories and Species, The Evolutionary and Cognitive Causes of the Species Problem. J. HEY. Oxford University Press. 2001. 217 pages. ISBN 0 19 514477 5. Price£ 34.95.(hardback). Genetics Research 79:107.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, and L. Z. Garamszegi. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome research 25:1656-1665.

Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society of London B: Biological Sciences 363:2997-3007.

Campagna, L., I. Gronau, L. F. Silveira, A. Siepel, and I. J. Lovette. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular ecology 24:4238-4251.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, and I. J. Lovette. 2017. Repeated divergent selection on pigmentation genes in a rapid finch radiation. Science Advances 3:e1602404.

Cariou, M., L. Duret, and S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3:846-852.

Chaves, J. A., E. A. Cooper, A. P. Hendry, J. Podos, L. F. De León, J. A. Raeymaekers, W. O. MacMillan, and J. A. C. Uy. 2016. Genomic variation at the tips of the adaptive radiation of Darwin's finches. Molecular ecology 25:5282-5295.

Chesser, R. T., K. J. Burns, C. Cicero, J. L. Dunn, A. W. Kratter, I. J. Lovette, P. C. Rasmussen, J. Remsen Jr, J. D. Rising, and D. F. Stotz. 2017. Fifty-eighth supplement to the American Ornithological Society's Check-list of North American Birds. The Auk 134:751-773.

Chui, C. K. S. and S. M. Doucet. 2009. A test of ecological and sexual selection hypotheses for geographical variation in coloration and morphology of golden-crowned kinglets (Regulus satrapa). Journal of Biogeography 36:1945-1957.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411-1423.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, and S. T. Sherry. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Darwin, C. 1859. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

de Queiroz, K. 2007. Species concepts and species delineation. Systematic Biology 56:879 - 886.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, and M. Hanna. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Anglais 43:491-498.

Dieckman, U. and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354-357.

Dobzhansky, T. 1935. A critique of the species concept in biology. Philosophy of Science 2:344-355.

Dobzhansky, T. 1950. Mendelian populations and their evolution. The American Naturalist 84:401-418.

Dobzhansky, T. G. 1937. Genetics and the origin of species. Columbia University Press, New York.

Doebeli, M. and U. Dieckmann. 2002. Speciation along environmental gradients.

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist:172-181.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for Ancient Admixture between Closely Related Populations. Molecular biology and evolution 28:2239-2252.

Eaton, D. A., A. L. Hipp, A. González-Rodríguez, and J. Cavender-Bares. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution 69:2587-2601.

Edwards, S. V., S. Potter, C. J. Schmitt, J. G. Bragg, and C. Moritz. 2016. Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proceedings of the National Academy of Sciences 113:8025-8032.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw, and C. M. Holzapfel. 2010. Resolving postglacial phylogeography

using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107:16196-16200.

Faircloth, B. C. and T. C. Glenn. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PloS one 7:e42543.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological Speciation: an integrative approach. Molecular ecology 23:513-521.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. Nature 336:61-64.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28:342-350.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35:124-138.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford Univ. Press, Oxford.

Forester, B. R., M. R. Jones, S. Joost, E. L. Landguth, and J. R. Lasky. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular ecology 25:104-120.

Freedman, A. H., I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han, P. M. Silva, M. Galaverni, Z. Fan, P. Marx, and B. Lorente-Galdos. 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS genetics 10:e1004016.

Frichot, E., S. D. Schoville, P. de Villemereuil, O. E. Gaggiotti, and O. François. 2015. Detecting adaptive evolution based on association with ecological gradients: orientation matters! Heredity 115:22-28.

Gavrilets, S. 2000. Waiting time to parapatric speciation. Proceedings of the Royal Society of London B: Biological Sciences 267:2483-2492.

Gavrilets, S. 2004. Fitness landscapes and the origin of species (MPB-41). Princeton, NJ: Princeton university press.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. Plos One 9:e90346-e90346.

Goldsmith, T. H. 1990. Optimization, Constraint, and History in the Evolution of Eyes. Quarterly Review of Biology 65:281-322.

Gordon, A. and G. Hannon. 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab. cshl. edu/fastx_toolkit.

Grant, B. R. and P. R. Grant. 2017. Watching speciation in action. Science 355:910-911.

Grant, P. R. 1986. Ecology and evolution of Darwin's finches. Princeton Univ. Press, Princeton, NJ.

Grant, P. R. 1998. Evolution on islands. Oxford University Press, Oxford, UK.

Grant, P. R. 2001. Reconstructing the evolution of birds on islands: 100 years of research. Oikos 92:385-403.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, and M. H.-Y. Fritz. 2010. A draft sequence of the Neandertal genome. science 328:710-722.

Griffith, S. C. and K. L. Buchanan. 2010. The zebra finch: the ultimate Australian supermodel. Emu 110:v-xii.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Anglais 43:1031-U1151.

Groth, J. G. 1993. Call matching and positive assortative mating in red crossbills. The Auk 110:398-401.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. Plos Genetics 5:e1000695-e1000695.

Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Research 27:1004-1015.

Harvey, M. G. and R. T. Brumfield. 2015. Genomic variation in a widespread Neotropical bird (Xenops minutus) reveals divergence, population expansion, and gene flow. Molecular phylogenetics and evolution 83:305-316.

Hey, J. 2001. Genes, categories, and species: The evolutionary and cognitive cause of the species problem. Oxford University Press.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics 6:e1000862-e1000862.

Irwin, K. 2010. A new and cryptic call type of the red crossbill. West Birds 41:10-25.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, and S. White. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55.

Jones, J. C., S. Fan, P. Franchini, M. Schartl, and A. Meyer. 2013. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Molecular ecology 22:2986-3001.

Kawecki, T. 2004. Genetic theories of sympatric speciation. Adaptive speciation. Cambridge University Press, Cambridge:36-53.

Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular ecology 22:2848-2863.

Korte, A. and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant methods 9:29.

Kraaijeveld, K., F. J. L. Kraaijeveld-Smit, and M. E. Maan. 2011. Sexual selection and speciation: the comparative evidence revisited. Biological Reviews 86:367-377.

Lamichhaney, S., J. Berglund, M. S. Almen, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerova, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371-375.

Lamichhaney, S., F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, and L. Andersson. 2016. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science 352:470-474.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the USA 78:3721-3725.

Lande, R. 1982. Rapid origin of sexual isolation and character divergence in a cline. Evolution 36:213-223.

Lawson, L. P. and K. Petren. 2017. The adaptive genomic landscape of beak morphology in Darwin's finches. Molecular Ecology.

Leggett, R. M., B. J. Clavijo, L. Clissold, M. D. Clark, and M. Caccamo. 2013. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics:btt702.

Lerner, H. R. L., M. Meyer, H. F. James, M. Hofreiter, and R. C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers. Current Biology 21:1838-1844.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. Barrett, J. G. Larson, J. D. Jensen, and H. E. Hoekstra. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312-1316.

Losos, J. B. and R. E. Ricklefs. 2009. Adaptation and diversification on islands. Nature 457:830-836.

Lovette, I. J., E. Bermingham, and R. E. Ricklefs. 2002. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. Proceedings of the Royal Society Biological Sciences Series B 269:37-42.

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, and Y. Liu. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.

Maia, R., C. M. Eliason, P.-P. Bitton, S. M. Doucet, and M. D. Shawkey. 2013. pavo: an R package for the analysis, visualization and organization of spectral data. Methods in Ecology and Evolution 4:906-913.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, and Z. Chen. 2005. Genome sequencing in open microfabricated high density picoliter reactors. Nature 437:376.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Research 23:1817-1828.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

Mayr, E. 1947. Ecological factors in speciation. Evolution:263-288.

Mayr, E. 1963. Animal species and evolution. Belknap Press, Cambridge, MA.

Mayr, E. 1995. Species, classification, and evolution. Biodiversity and Evolution. National Science Museum Foundation, Tokyo:3-122.

Mayr, E. and P. D. Ashlock. 1969. Principles of systematic zoology.

McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66:526-538.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20:1297-1303.

Milá, B., P. Aleixandre, S. Alvarez-Nordström, J. McCormack, E. Ketterson, and J. Atwell. 2016. More than meets the eye: Lineage diversity and evolutionary history of Dark-eyed and Yellow-eyed juncos. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco (ED Ketterson and JW Atwell, Editors). University of Chicago Press, Chicago, Illinois, USA:179-198.

Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society B-Biological Sciences 274:2653-2660.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Moyle, R. G., C. E. Filardi, J. J. Smith, and J. M. Diamond. 2009. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". Proceedings of the National Academy of Sciences of the United States of America 106:1863-1868.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences 367:343-353.

Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus, S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Nosil, P. 2012. Ecological Speciation. Oxford University Press.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:“isolation by adaptation” and multiple roles for divergent selection. Evolution 62:316-336.

Nosil, P., J. L. Feder, S. M. Flaxman, and Z. Gompert. 2017. Tipping points in the dynamics of speciation. Nature Ecology & Evolution 1:0001.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends in Ecology and Evolution 16:325-413.

Parchman, T. L., C. A. Buerkle, V. Soria-Carrasco, and C. W. Benkman. 2016. Genome divergence and diversification within a geographic mosaic of coevolution. Molecular ecology 25:5705-5718.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Price, T. 1987. Diet variation in a population of Darwin's finches. Ecology 68:1015-1028.

Price, T. 2008. Speciation in birds. Roberts and Company, Greenwood Village, Colorado.

Queiroz, K. and M. J. Donoghue. 1988. Phylogenetic systematics and the species problem. Cladistics 4:317-338.

R_Core_Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R_Studio_Team. 2015. RStudio: Integrated Development for R. R Studio, Inc., Boston, MA.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24:4348-4370.

Rheindt, F. E., M. K. Fujita, P. R. Wilton, and S. V. Edwards. 2013. Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. Systematic Biology 63:134-152.

Rice, A. M., A. Rudh, H. Ellegren, and A. Qvarnström. 2011. A guide to the genomics of ecological speciation in natural animal populations. Ecology Letters 14:9-18.

Riesch, R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K. Lucek, E. Hellen, V. Soria-Carrasco, and S. R. Dennis. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. Nature Ecology & Evolution 1:0082.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala: revisionary studies and comparative geography. Bulletin of the AMNH; v. 162, article 5.

Ruegg, K., E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Molecular Ecology 23:4757-4769.

Rundle, H. D. and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336-352.

Sato, A., H. Tichy, C. O'HUigin, P. R. Grant, B. R. Grant, and J. Klein. 2001. On the Origin of Darwin's Finches. Molecular Biology and Evolution 18:299-311.

Schiffels, S. and R. Durbin. 2014. Inferring human population size and separation history from multiple genome sequences. Anglais 46:919-925.

Schluter, D. 1996. Ecological causes of adaptive radiation. The American Naturalis 148:S40-S64.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology and Evolution 16:372-380.

Schmieder, R. and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864.

Schmitz-Ornes, A. 2006. Using colour spectral data in studies of geographic variation and taxonomy of birds: examples with two hummingbird genera, Anthracothorax and Eulampis. Journal of Ornithology 147:495-503.

Scholes, E. 2008. Evolution of the courtship phenotype in the bird of paradise genus Parotia (Aves : Paradisaeidae): homology, phylogeny, and modularity. Biological Journal of the Linnean Society 94:491-504.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, and e. al. 2014. Genomics and the origin of species. Nature Reviews Genetics 15:176-192.

Seutin, G. 1991. Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology 69:82-90.

Simpson, G. 1953. The major features of evolution. and G. Simpson, Tempo and Mode in:143.

Simpson, G. G. 1961. Principles of animal taxonomy. Columbia University Press, New York.

Sorenson, M. D., K. M. Sefc, and R. B. Payne. 2003. Speciation by host switch in brood parasitic indigobirds. Nature 424:928.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014. Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. Science 344:738-742.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smajda, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology and Evolution 25:705-712.

Stinchcombe, J. and H. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158.

Stoddard, M. C. and R. O. Prum. 2008. Evolution of avian plumage color in a tetrahedral color space: A phylogenetic analysis of new world buntings. American Naturalist 171:755-776.

Sullivan, K. A. 1999. Yellow-eyed junco (*Junco phaeonotus*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America Inc., Philadelphia, PA.

Szulkin, M., P. A. Gagnaire, N. Bierne, and A. Charmantier. 2016. Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. Molecular ecology 25:542-558.

Templeton, A. R. 1989. The meaning of species and speciation: a genetic perspective. The units of evolution: Essays on the nature of species:159-183.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. Taxon:233-239.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends in ecology & evolution 16:381-390.

Villemereuil, P. and O. E. Gaggiotti. 2015. A new FST-based method to uncover local adaptation using environmental variables. Methods in Ecology and Evolution 6:1248-1258.

Wagner, C., L. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366-369.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22:787-798.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, and S. Fairley. 2010. The genome of a songbird. Nature 464:757.

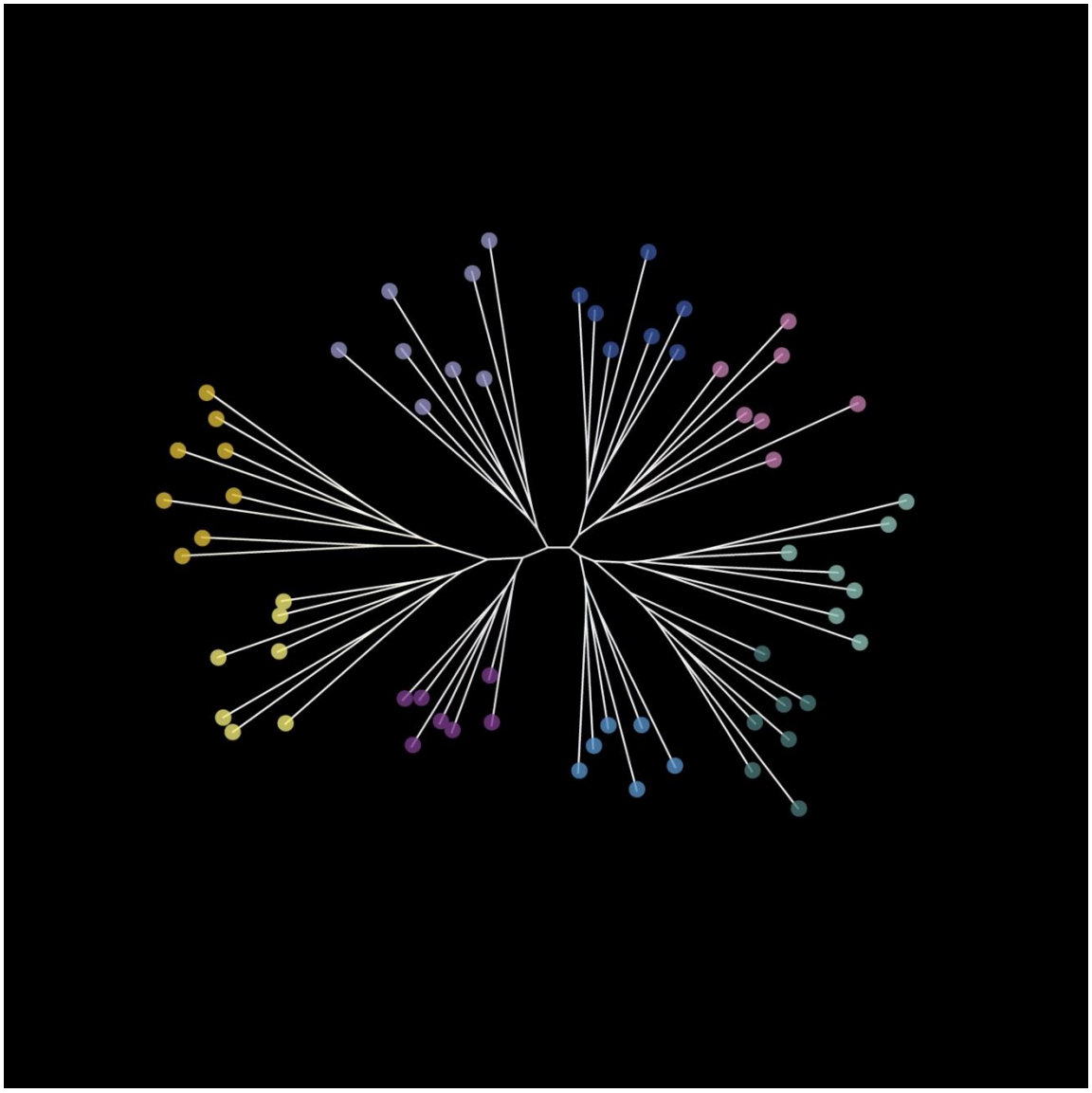
West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. The Quarterly Review of Biology 58:155-183.

Wiley, E. O. 1978. The evolutionary species concept reconsidered. Systematic zoology 27:17-26.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the national Academy of sciences 106:13875-13879.

Zamudio, K. R., R. C. Bell, and N. A. Mason. 2016. Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. Proceedings of the National Academy of Sciences 113:8041-8048.

CHAPTER I: Rapid postglacial diversification and long-term stasis within the songbird genus Junco: phylogeographic and phylogenomic evidence



Guillermo Friis, Pau Aleixandre, Ricardo Rodríguez-Estrella, Adolfo G. Navarro-Sigüenza, Borja Milá

Published in Molecular Ecology in 2016

Abstract

Natural systems composed of closely related taxa that vary in the degree of phenotypic divergence and geographic isolation, provide an opportunity to investigate the rate of phenotypic diversification and the relative roles of selection and drift in driving lineage formation. The genus *Junco* (Aves: Emberizidae) of North America includes parapatric northern forms that are markedly divergent in plumage pattern and color, in contrast to geographically isolated southern populations in remote areas that show moderate phenotypic divergence. Here, we quantify patterns of phenotypic divergence in morphology and plumage color, and use mitochondrial DNA genes, a nuclear intron, and genome-wide SNPs to reconstruct the demographic and evolutionary history of the genus to infer relative rates of evolutionary divergence among lineages. We found that geographically isolated populations have evolved independently for hundreds of thousands of years despite little differentiation in phenotype, in sharp contrast to phenotypically diverse northern forms, which have diversified within the last few thousand years as a result of the rapid postglacial recolonization of North America. SNP data resolved young northern lineages into reciprocally monophyletic lineages, indicating low rates of gene flow even among closely related parapatric forms, and suggesting a role for strong genetic drift or multifarious selection acting on multiple loci in driving lineage divergence. Juncos represent a compelling example of speciation-in-action, where the combined effects of historical and selective factors have produced one of the fastest cases of speciation known in vertebrates.

Key words: phylogeography, phylogenomics, rapid speciation, postglacial expansion, avian radiation, GBS

Introduction

One of the most remarkable patterns in nature is the difference in the degree and rate of phenotypic diversification among clades and geographic areas. At large macroevolutionary scales, morphological diversification shows marked variation across clades, and it has been suggested to be more strongly related to the number of species in the clade than to its age (Rabosky et al. 2012; Rabosky et al. 2013). The fact that species number is also decoupled from clade age, suggests the role of ecological limits to diversification and the interplay of a number of factors in determining patterns of species richness, including latitude, climate, ecology and biogeography (Currie et al. 2004; Phillimore et al. 2007; Rabosky 2009; Mahler et al. 2010; Yoder et al. 2010; Title and Burns 2015). In addition, different clades have shown different degrees of 'evolvability', or the capacity to modify their morphology in the face of strong selection, as exemplified in birds by the striking diversification of Hawaiian honeycreepers, ten times greater than that of the Hawaiian thrushes, despite having colonized the archipelago at similar times (Lovette *et al.* 2002). In general, however, uncertainty concerning past extinction events, branch lengths and ancestral trait values at large phylogenetic scales complicates efforts to distinguish between time-dependent and speciation-dependent processes (Ricklefs 2006). Clearly, understanding the factors responsible for patterns of species richness among clades and the evolvability of different lineages requires not only improved macroevolutionary approaches, but also more detailed studies at finer phylogenetic scales.

Taxonomic groups at inter- and intra-specific levels that cover a broad range of spatial scales and ecological conditions allow a closer study of the relative roles of selection and drift in driving evolutionary divergence, and the rate of phenotypic diversification and speciation. Specifically, ideal scenarios for studying speciation mechanisms are provided by those natural systems that (i) are composed of recently evolved and closely related intra and inter-specific lineages or populations, minimizing the probability of extinction; (ii) are found at different spatial scales, from local to regional, and covering a range of ecological conditions; (iii) differ prominently in the degree of geographic isolation, from none to moderate gene flow; and (iv) show marked differentiation in phenotypic traits related to fitness.

The songbird genus *Junco* of North and Central America provides such a system, considering its striking patterns of geographic variation spanning the entire continuum from the local population to the species, and from complete isolation to freely interbreeding parapatric forms (Miller 1941; Nolan et al. 2002; Fig. 1.1A; see 'Study System' from the General Introduction).

Here, we use an extensive taxonomic sampling of the genus that includes all major phenotypic forms, and we reconstruct the evolutionary and demographic history of the complex to assess differences in patterns and rates of phenotypic diversification among *Junco* lineages. We also explore the relative roles of historical factors (geography, shifts in distribution range, and time in allopatry) and selective factors (ecology and sexual selection), in explaining differences in phenotypic diversification among clades. The study by Milá *et al.* (2007; see 'Study System' from General Introduction) used spatial patterns of mitochondrial DNA (mtDNA) control region diversity to show that dark-eyed junco forms presented genetic signatures of recent population expansion coincident with the aftermath of the Last Glacial Maximum (LGM) about 18,000 years ago. This was in contrast to lineages in Guatemala (*alticola*), and Guadalupe Island (*insularis*), which despite their lower divergence in plumage color, turned out to be an order of magnitude older (Milá et al. 2007; Aleixandre et al. 2013), suggesting that different rates of phenotypic diversification have operated in different clades and different geographic areas. Here, in addition to mtDNA markers, we use a large set of genome-wide single-nucleotide polymorphism (SNP) loci to achieve the resolution needed to reconstruct the evolutionary history of the group, and to test the dark-eyed junco post-glacial diversification hypothesis from a genomic perspective. We also use morphometric and colorimetric datasets to quantify phenotypic differences among forms, assess correlations between phenotypic and genetic distances, and infer the contribution of natural and sexual selection in the radiation by assessing the relative divergence of morphological traits, and plumage coloration, respectively.

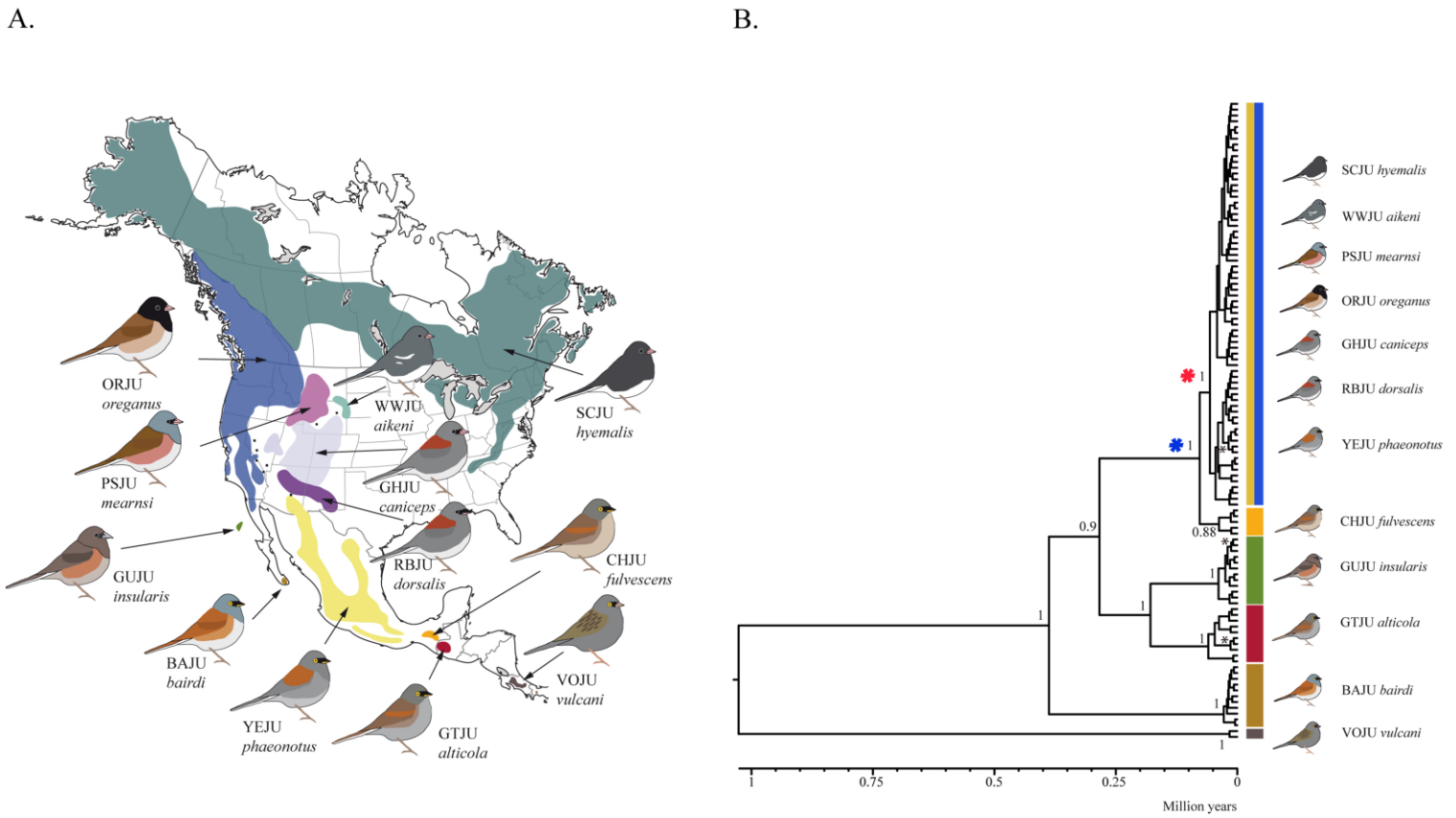


Figure 1.1. Geographic distribution of phenotypic variation and phylogenetic relationships among *Junco* forms. A. Distribution map of the different junco forms. Colored areas correspond to the breeding ranges of the major forms (see Table 1 for a detailed nomenclature). Dots represent isolated localities with hybrid/inter-mediate individuals. B. Ultrametric Bayesian phylogeny based on four mitochondrial regions (CR, ND2, COI and ATPase) and the FGB-I5 intron. Posterior probability values are shown at each node. Black asterisks correspond to additional nodes with support values above 0.9. The node separating *fulvescens* and the clade of recently radiated northern forms (blue asterisk) was timed at 79,600 years before present (95% HPD: 41,300-121,800) using BEAST, while the TMRCA of the northern form clade along (red asterisk) was estimated at 58,300 years ago (95%HPD: 31,300-88,600) in BEAST, and 16,132 years ago (95%HPD: 6,874-36,335) in IMa2. Colors correspond to those used in Fig. 1.1A.

Materials and methods

Sequencing of mitochondrial and nuclear markers

We amplified four regions of the mtDNA from 212 samples distributed among all major junco forms: 349 base pairs (bp) of the hypervariable region I of the control region (CR) using primers LGL2 (L2263, 5′-GGCCACATCAGACAGTCCAT-3′) and H417 (H2607, 5′-AGTAGCTCGGTTCTCGTGAG-3′) (Tarr 1995). L16150 (5'-CCTCYAYCWCCARCTCCCAAAGC-3') (Sorenson et al. 1999) was used instead for problematic samples; 640 bp of the cytochrome c oxidase I (COI) gene using primers "BirdF1" (5′-TTCTCCAACCACAAAGACATT-3′) and "BirdR1" (5′-CGTGGGAGATAATTCCAAATCCTG-3′) (Hebert et al. 2004). "BirdR2" (5'-ACTACATGTGAGATGATTCCGAATCCAG-3') was used to replace "BirdR1" to amplify samples from Guadalupe island; 985 bp of the NADH dehydrogenase subunit 2 (ND2) using primers L5215 (5'-TATCGGGCCCATACCCCGAAAAT-3') (Hackett 1996) and H6313 (5'-CTCTTATTTAAGGCTTTGAAGGC-3') (Sorenson et al. 1999) and 870 bp of the ATPase genes 8 & 6 (ATPase) using primers L8929 (5'-GGACAATGCTCAGAAATCTCGCGG-3') (Eberhard and Bermingham 2005) and H9855 (5'-ACGTAGGCTTGGATTATKGCTACWGC-3') (Sorenson et al. 1999). We also amplified the nuclear fibrinogen beta chain (FGB) gene, intron 5, using primers Fib5 (5'-CGCCATACAGAGTATACTGTGACAT-3') and Fib6 (5'-GCCATCCTGGCGATTCTGAA-3') (Kimball et al. 2009) for a subdataset of 110 individuals (Table 1). PCR products were purified with an ethanol precipitation and sequenced in an ABI 3730X automated sequencer. The amplification of nuclear copies of the control region is unlikely since we amplified it together with a coding fragment that was unambiguously translated into its aminoacid sequence. Sequences were aligned using Sequencher 4.1.4 (GeneCodes) and the accuracy of variable sites was checked visually on the individual chromatograms. Sequence alignments for each marker were imported to DnaSP 5.10.01 (Librado & Rozas 2009) and haplotype data files were generated for each of them independently. Nuclear intron sequences were phased using 1000 MCMC iterations as implemented by the program PHASE in the DnaSP package.

Genotyping of genome-wide SNP loci

The present chapter is based in the 95 samples genotyped in the Plate I (see General Methods) belonging to the following taxa (with sample sizes in parentheses): *hyemalis* (7 individuals from the subspecific *carolinensis* form), *aikeni* (7), *oreganus* (8 *thurberi* and 7 *montanus* individuals), *mearnsi* (7), *caniceps* (8), *dorsalis* (8), *phaeonotus* (7 *palliatus* and 8 *phaeonotus* individuals), *fulvescens* (3), *alticola* (7), *insularis* (7), *bairdi* (7), and *vulcani* (2) (Table 1.1; see Table 1 from the General Introduction for a detailed nomenclature). Sequencing of 2 of the 95 samples failed (1 *thurberi* and 1 *montanus*). Here we used the GBS analysis pipeline version 3.0.160 (Glaubitz *et al.* 2014), an extension to the Java program TASSEL (Bradbury *et al.* 2007), to call the SNPs from the sequenced GBS library. Because the draft genome of junco was not yet available for the analyses included in this first Chapter (see 'Whole genome sequencing and assembly' from General Methods), we used as a reference the *Zonotrichia albicollis* genome [GenBank accession: ARWJ00000000] (Romanov et al. 2011). To account for genotyping errors, the TASSEL3-GBS pipeline excludes individual tags represented by less than five reads before genotyping. During variant discovery, loci for which less than 10% of the taxa had one or more tags were not used, and SNPs with a minor allele frequency (MAF) below 0.01 were excluded. Average sequencing error rate used to decide between heterozygous and homozygous calls was set by default to 0.1. We set a minimum sample call rate for each SNP of 0.8 and a minimum SNP call rate for each sample of 0.1. The resulting data matrix included 258,933 SNPs for 93 samples.

Phylogenetic analysis and molecular dating

We constructed a phylogeny using the concatenated mtDNA markers and the FGB intron 5, for a total of 110 individuals. The sequences were analyzed with Partition Finder (Lanfear et al. 2012) which found statistical support for partitioning by gene but not by codon position. We then used jModelTest (Posada 2008) for determining the model of molecular evolution of each marker following the Bayesian Information Criterion (BIC), which were Hasegawa-Kishino-Yano with gamma distributed rate of among-site variation (HKY + G) for COI, ND2 and CR; Tamura-Nei (TrN) + G for ATPase; and HKY for the FGB intron. We simultaneously estimated tree topology and node divergence times through Bayesian inference using the program BEAST

1.8.1 (Drummond et al. 2012). A preliminary run using a lognormal relaxed clock model for all markers yielded moderate to low deviation from clock-like behavior for ATPase, CR and especially COI (Coefficient of variation = 0.185 as checked in TRACER v.1.6 (Rambaut et al. 2014), while ND2 and FGB intron 5 showed high substitution rate variability across the tree. Following Lerner *et al.* (2011), we fixed an evolutionary rate of 0.016 substitutions per site per million years for the COI gene under a strict clock model, while estimating the rates under a strict clock model for ATPase and CR, and a lognormal relaxed clock model for ND2 and FGB. A coalescent model was chosen over a birth-death model based on the path sampling (PS) and stepping stone (SS) sampling methods which have been shown to outperform the harmonic mean and Akaike's information criterion (AICM) to compute marginal likelihoods (Baele et al. 2012). Since we do not know the precise demographic history of the genus, we implemented the Bayesian Skyline coalescent model as available in BEAST, which can fit a wide range of demographic scenarios while taking into account phylogenetic uncertainty (Drummond et al. 2005). We set the number of groups to six based on the number of main lineages in the haplotype network (see Results) and used a piecewise-linear Skyline model. We ran the analysis for 400M iterations, sampling the chains every 8,000 steps. Convergence and effective sample sizes (ESS) were checked with TRACER v.1.6. We also constructed haplotype networks for all four mtDNA markers (COI, ATPase, ND2 and CR) using the median-joining algorithm (Bandelt et al. 1999) in the program Network 4.6.1.2 (fluxus-engineering.com). As an alternative method for dating the postglacial expansion, we also estimated the time to the most recent common ancestor (TMRCA) of the young yellow-eyed and dark-eyed forms using the COI evolutionary rate in IMa2 (Hey 2010), and applying a pure isolation model that excluded migration. Based on preliminary runs, we set priors for maximum time of population splitting and maximum population size of 20 and 60, respectively. Generation time was set to 2 years since juncos do not usually breed successfully their first year. We run 40 chains of 30 million steps each, sampling every 100 steps after a 100,000-step burn-in, with heating in a geometric increment model with a term 1 of 0.96 and a term 2 of 0.9.

We calculated Nei's (Nei 1987) unbiased haplotypic diversity (h) and nucleotide diversity (π) indices for all lineages using the four mitochondrial markers in Arlequin 3.5.1.2 (Excoffier et al. 2005). To test for past sudden changes in population size we used Fu's test of neutrality, which detects departures from neutrality in scenarios with biased frequencies of rare alleles or young mutations in non-recombinant sequences (Fu 1997). We used Arlequin to obtain Fu's F_S values, and interpreted large and significant negative values as an excess of recent mutations caused by a recent population expansion. We also used Arlequin to compute the average number of pairwise differences corrected for intragroup variation for the ND2 mitochondrial marker.

Table 1.1. Number of genotyped individuals per junco form and molecular marker. State abbreviations are the following: Virginia (VA), Wyoming (WY), Alaska (AK), California (CA), Idaho (ID), Utah (UT), and Arizona (AZ) in the USA; British Columbia (BC) in Canada; Coahuila (COAH), Mexico City (CM), Durango (DGO), Michoacán (MICH), Nuevo León (NL), Oaxaca (OAX), and Chiapas (CHIS) in Mexico; and Huehuetenango (HUE) in Guatemala.

Group	*State*	*mtDNA*	*FGB-I5*	*SNPs*
hyemalis	VA	6	16	7
aikeni	WY	8	6	7
oreganus	AK, BC, CA, ID	36	18	15
mearnsi	WY	8	3	7
caniceps	UT	9	4	8
dorsalis	AZ	10	4	8
phaeonotus	AZ, COAH, CM, DGO, MICH, NL, OAX	76	19	15
fulvescens	CHIS	8	5	3
insularis	Guadalupe Is.	23	12	7
bairdi	BCS	11	11	8
alticola	HUE	15	10	8
vulcani	Costa Rica	2	2	2
Total		212	110	95

Genome-wide structure from SNP data

To explore genome-wide population structure among recently diverged junco taxa and assess its correspondence with phenotypic groups, we ran a principal components analysis (PCA) using SNP data. We constructed a data matrix excluding older lineages (*vulcani*, *bairdi*, *insularis*, *alticola* and *fulvescens*) as well as samples with a proportion of genotyped variable sites below 0.6, or with more than 20% heterozygous positions. Variant sites missing in more than 85% of the samples or presenting a minor allele frequency (MAF) below 0.03 were also removed. We filtered out SNPs putatively under selection using BayeScan (Foll and Gaggiotti 2008). BayeScan computes and decomposes per-SNP F_{ST} scores into a population-specific component shared by all loci that approximates population related effects, as well as a locus-specific component shared by all populations, which accounts for selection. Then it compares two models of divergence, with and without selection, and assumes a departure from neutrality when the locus-specific component is necessary to explain a given diversity pattern (Foll 2012). Original *vcf* files were converted to BayeScan format using PGDSpider version 2.0.5.1 (Lischer and Excoffier 2012). We used default options in BayeScan and set the thinning interval size to 100 to ensure convergence. Under these conditions, the neutral model has a prior likelihood 10 times higher than the model with selection at any given locus. We defined junco populations according to phenotype (see Table 1 from the General Introduction), and for each SNP we obtained the posterior probability for the selection model and the F_{ST} coefficient averaged over populations. For outlier detection, we implemented a false discovery rate of 0.1. From a starting dataset of 101,743 loci, 597 significant outliers were excluded. Finally, to filter out the SNPs under linkage disequilibrium (LD) we used the function snpgdsLDpruning from the package SNPrelate (Zheng 2012) in R version 3.2.2 (R Core Team 2013). We applied the correlation coefficient method with a threshold of 0.2 (method ="corr", ld.threshold=0.2), resulting in a final data matrix of 24,832 SNPs. We then used the function snpgdsPCA also available in SNPrelate to perform the PCA and obtain the eigenvectors to be plotted.

Phylogenomic analyses with SNP data

To reconstruct phylogenetic relationships among junco forms using SNP data, we constructed two different data matrices, one including all taxa except the divergent volcano junco, and the other including the northern forms only (excluding older lineages *bairdi*, *insularis*, *alticola* and *fulvescens*). We applied the same quality filtering controls than in the construction of the data matrix for the PCA (see above), except that the value for the threshold for missing genotyped sites was raised to 0.8 for the full-genus matrix. To fulfill neutrality assumptions for phylogenetic analyses, we filtered out SNPs putatively under selection from each matrix using the software BayeScan with the same settings used for the PCA. Once outliers were excluded, we created different matrices corresponding to different slices of the distribution of per-locus global-F_{ST} values among forms (Keller et al. 2013; Puebla et al. 2014). We tested the phylogenetic resolution of neutral SNPs using three different F_{ST} intervals: (i) 100% of the SNP dataset, (ii) top 25% F_{ST} SNPs, and (iii) top 10% F_{ST} SNPs (Table 1.2).

We constructed maximum likelihood phylogenies with RAxML 8.1.11 (Stamatakis 2014) in the CIPRES Science Gateway (Miller et al. 2010). We implemented a General Time Reversible (GTR) + gamma model of sequence evolution and applied the Lewis ascertainment bias correction (Stamatakis 2014) to the full genus tree which improved branch support. We used the rapid bootstrap algorithm (Stamatakis et al. 2008) with 100 replicates to compute a support value for the topology. We repeated the northern radiation phylogenetic reconstruction for the northern types including three samples of *fulvescens*, from Chiapas, so we could root the tree and discuss the evolutionary history of the group. Finally, to assess the significance of the phylogenetic signal in our datasets, we replicated the phylogenetic analysis on a matrix of randomized individuals assigned stochastically to an equal number of taxa, and using the top 25% F_{ST} SNPs to generate a maximum likelihood tree (Table 1.2).

To compare genetic and phenotypic distances, we computed a matrix of pairwise Nei's distances from the entire SNP dataset after filtering for selection using the package *adegenet* (Jombart 2008) in R version 3.2.2.

Table 1.2. Data matrices used in phylogenomic analysis. Shown for each matrix are number of samples, number of SNP loci and F_{ST} ranges for datasets used in phylogenomic reconstructions.

Dataset	Sample N	F_{ST} percentile	SNPs number	F_{ST} range
Full genus (except *vulcani)*	79	100%	130,700	0.085-0.371
		Top 25%	32,674	0.180-0.371
		Top 10%	13,070	0.186-0.371
Northern forms	64	100%	101,145	0.053-0.205
		Top 25%	25,286	0.063-0.205
		Top 10%	10,114	0.067-0.205
Northern forms and *fulvescens*	67	100%	105,572	0.061-0.258
		Top 25%	26,393	0.086-0.258
		Top 10%	10,557	0.090-0.258
Randomized dataset	64	Top 25%	25,387	0.036-0.150

Phenotypic data and analysis

We examined overall morphological and colorimetric differentiation (see General Methods for a detailed explanation of the used variables) among junco forms using male data from museum specimens (Table 1.3) in two different discriminant function analyses (DFA). Morphological variables were previously transformed using natural logarithms. Analyses were carried out with SPSS 17.0 and R version 3.2.2.

Phenotypic versus genetic distance

To compare genetic and phenotypic distances we used pairwise Nei's distances based on SNPs (see "*Phylogenomic analysis with SNP data*") along with Euclidean distances among group centroids from the DFAs of phenotypic data computed in SPSS 17.0. All distance values were rescaled to 1 and graphed in two heat-map color matrices using the function 'image' from the R-package *graphics*. In these heat-map

Table 1.3. Sample sizes and localities of specimens from museum collections used for measurements of morphology and color. Specimens were measured at the following museums: San Diego Natural History Museum (SDMNH), Los Angeles Museum of Natural History (LAMNH), The Museum of Vertebrate Zoology (MVZ), The Moore Laboratory of Zoology at Occidental College (MLZ), The National Museum of Natural History (NMNH) and The American Museum of Natural History (AMNH). See Appendix I for catalog numbers.

Group	*Locality*	*n males*	*n females*
hyemalis	Alberta, Canada	5	6
	Newfoundland, Canada	3	3
	Northwest Territories, Canada	0	3
	Nova Scotia, Canada	2	1
	Québec, Canada	1	1
	Alaska, USA	12	7
	Georgia, USA	2	1
	Kentucky, USA	1	1
	Maryland, USA	1	0
	Maine, USA	1	0
	North Carolina, USA	1	3
	New Hampshire, USA	1	0
	Tennessee, USA	2	2
	Virginia, USA	3	4
	West Virginia, USA	2	0
aikeni	New Brunswick, Canada	2	2
	Colorado , USA	3	5
	Montana, USA	8	6
	South Dakota, USA	1	0
oreganus	British Columbia, Canada	7	6
	Alaska, USA	11	7
	Arizona, USA	0	1
	California, USA	49	43
	Oregon, USA	8	6
	Washington, USA	1	0
	Baja California, Mexico	28	21
mearnsi	Arizona, USA	5	4
	Idaho, USA	0	3

	Montana, USA	17	15
	Wyoming, USA	4	0
caniceps	Colorado, USA	5	2
	New Mexico, USA	9	2
	Nevada, USA	5	6
	Utah, USA	0	1
dorsalis	Arizona, USA	12	6
	New Mexico, USA	3	7
phaeonotus	Arizona, USA	10	10
	Chihuahua, Mexico	7	7
	Durango, Mexico	1	2
	Guerrero, Mexico	8	5
	Mexico DF, Mexico	5	9
	Oaxaca, Mexico	2	3
	Veracruz, Mexico	2	0
fulvescens	Chiapas, Mexico	23	12
insularis	Isla Guadalupe, Mexico	18	12
bairdi	Baja California, Mexico	21	15
alticola	Guatemala, Guatemala	23	16
	Chiapas, Mexico	0	2
vulcani	Costa Rica	19	17
TOTAL		**354**	**285**

plots, the upper hemi-matrix depicts pairwise distances for phenotypic traits (morphology or color), and the lower hemi-matrix shows pairwise genetic distances based on SNPs. We also tested the correlation between genetic and Euclidean distances from morphometric and colorimetric traits by means of a Mantel test in the R-package *vegan* (Oksanen et al. 2005).

Results

Phylogeography and divergence times

Analysis of mtDNA and intron sequences revealed the existence of six well-supported phylogenetic clades corresponding, respectively, to junco populations in Costa Rica (*vulcani*), the tip of Baja California (*bairdi*), Guatemala (*alticola*), Guadalupe Island (*insularis*), Chiapas in southern Mexico (*fulvescens*), and a large clade composed of all remaining taxa from mainland Mexico, USA and Canada (Fig. 1.1B, Table 1.4). The form *bairdi* appears as the most basal lineage in the yellow-eyed/dark-eyed complex, with the clade including the rest of taxa divided into two major subclades, one formed by *alticola* and *insularis*, and the other by *fulvescens* and a single shallow clade containing all northern taxa. Molecular dating of the split between *fulvescens* and the northern clade yielded a mean value of 79,600 years before present [95% Highest Posterior Density interval (HPD): 41,300-121,800] while the TMRCA for the northern clade itself resulted in a mean value of 58,300 years (95% HPD: 31,300-88,600) when estimated with BEAST 1.8.1, and 16,132 years before present (95% HPD: 6,874-36,335) when using IMa2.

Haplotype networks for each of the mtDNA markers show topologies that are largely consistent with that of the Bayesian phylogeny (Fig. 1.2), although careful inspection reveals several interesting differences among genes Thus, the ND2 gene provides clear separation of the older lineages (*vulcani*, *bairdi*, *insularis* and *alticola*) yet mainland yellow-eyed and dark-eyed junco forms are clustered into a single haplogroup (Fig. 1.2A). In contrast, ATPase genes show relatively less differentiation between some isolated lineages like *insularis* and *alticola*, but instead separate mainland yellow-eyed juncos (*phaeonotus*) from all dark-eyed junco forms to the north (Fig. 1.2B). The COI gene shows a similar pattern to that of ND2, with more marked divergence of isolated clades (Fig. 1.2C), and the CR, the only non-coding region used, is the least informative marker at the lineage level, likely due to homoplasy resulting from its high mutation rate (Fig. 1.2D). A consistent sign is the presence of a "star-like" distribution of haplotypes in the north, with a high-frequency haplotype surrounded by several closely-related, low-frequency haplotypes, a pattern associated with a recent population expansion (Hewitt 1996).

Table 1.4. Genetic distances among junco forms. Nei's distances based on SNP data from genotyping-by-sequencing (above diagonal) and the average number of pairwise changes among forms corrected for intragroup variation for the ND2 gene (below diagonal, with intra-population divergence values along diagonal, in bold italics).

	hyemalis	*aikeni*	*mearnsi*	*oreganus*	*thurberi*	*montanus*	*caniceps*	*dorsalis*	*palliatus*	*phaeonotus*	*fulvescens*	*alticola*	*insularis*	*bairdi*	*vulcani*
hyemalis	***0.933***	0.079	0.075	-	0.063	0.071	0.079	0.064	0.089	0.112	0.160	0.129	0.182	0.250	-
aikeni	0.033	***0.000***	0.074	-	0.069	0.073	0.079	0.074	0.084	0.102	0.168	0.130	0.180	0.252	-
mearnsi	0.033	0.000	***2.000***	-	0.064	0.070	0.077	0.066	0.083	0.104	0.158	0.124	0.179	0.247	-
oreganus	0.037	0.004	0.004	***0.514***	-	-	-	-	-	-	-	-	-	-	-
thurberi	-	-	-	-	-	0.061	0.068	0.054	0.075	0.100	0.147	0.116	0.170	0.235	-
montanus	0.006	0.000	0.000	-0.011	-	***0.500***	0.073	0.064	0.077	0.098	0.155	0.120	0.172	0.242	-
caniceps	0.033	0.000	-0.028	0.004	-	0.000	***0.222***	0.068	0.081	0.100	0.160	0.125	0.179	0.248	-
dorsalis	0.033	0.000	0.000	0.004	-	0.000	0.000	***0.200***	0.070	0.097	0.132	0.108	0.177	0.238	-
palliatus	0.322	0.289	0.239	0.293	-	0.289	0.289	0.289	***2.108***	0.084	0.152	0.113	0.181	0.252	-
phaeonotus	0.029	0.004	-0.009	0.008	-	0.000	0.004	0.004	0.293	***0.383***	0.175	0.130	0.200	0.274	-
fulvescens	1.033	1.000	1.000	1.004	-	0.833	1.000	1.000	1.289	1.004	***0.000***	0.120	0.268	0.326	-
alticola	13.290	13.257	13.257	13.261	-	13.090	13.257	13.257	13.432	13.261	12.257	***0.686***	0.229	0.293	-
insularis	15.458	15.439	15.439	15.443	-	15.272	15.439	15.439	15.618	15.394	14.526	4.592	***1.644***	0.271	-
bairdi	15.033	15.000	15.000	14.917	-	14.833	15.000	15.000	15.175	15.004	14.000	7.057	9.395	***0.182***	-
vulcani	29.033	29.000	29.000	29.004	-	28.833	29.000	29.000	29.175	28.955	28.000	24.924	25.308	28.000	***0.000***

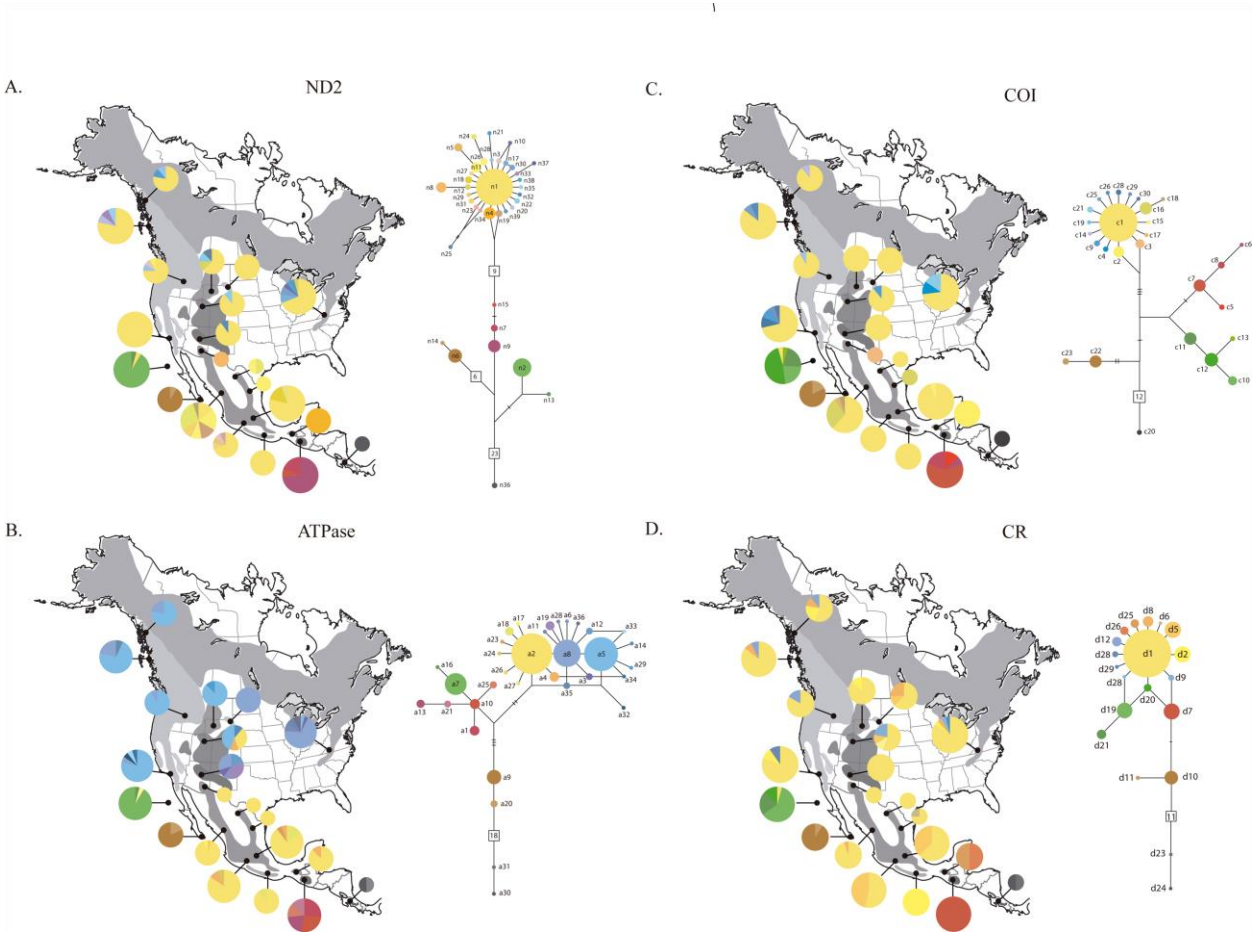


Figure 1.2. MtDNA structure and diversity across the junco range. Breeding range (gray areas), sampling sites (black dots), mtDNA haplotype frequency per locality (pie charts) and median-joining haplotype networks for ND2 (A), ATPase (B), COI (C) and he control region's hypervariable region I. On the maps, pie chart size is proportional to the number of samples, and colors correspond to the different haplotypes in the population depicted in the networks on the right, which also correspond to colors used in Fig. 1.1. In the networks, haplotypes are represented by circles sized in proportion to their frequency in the population. Each branch represents a single nucleotide change, with additional mutations indicated by bars along branches. Numbers in squares indicate nucleotide changes along longer branches.

Genetic diversity and historical demography

Sequencing of the four mitochondrial regions for 212 individuals produced 24 haplotypes for COI (640 bp), 28 haplotypes for ATPase (870 bp), 34 haplotypes for ND2 (985 bp), and 19 haplotypes for the non-coding CR (349 bp); resulting in 85 different haplotypes for the concatenated sequence (2844 bp). Indices of haplotype and nucleotide diversity were generally high (above 0.8 and 0.001 for *h* and π, respectively) except for three particularly small and isolated populations: *aikeni* (*h* = 0.607; π = 0.0002), *bairdi* (*h* = 0.473; π = 0.0004) and *fulvescens* (*h* = 0.679; π = 0.0002) (Table 5). Three of the taxa (*hyemalis*, *oreganus* and *phaeonotus*) showed evidence of population expansions as determined by significantly negative F_s values (Table 1.5).

Table 1.5. Genetic diversity and demographic history of junco forms. Indices are based on concatenated mitochondrial markers (CR, COI, ND2 and ATPase) and include haplotype diversity (*h*), nucleotide diversity (π) and Fu's test of population expansion (F_s). Asterisks correspond to $P \leq 0.02$.

Taxa	*n*	*No. haps.*	*H*	π	F_s
hyemalis	6	6	1.000 +/- 0.096	0.0012 +/- 0.0008	-3.03*
aikeni	8	3	0.607 +/- 0.164	0.0002 +/- 0.0002	-0.48
oreganus	36	17	0.800 +/- 0.068	0.0008 +/- 0.0005	-11.41*
mearnsi	8	6	0.893 +/- 0.111	0.0010 +/- 0.0007	-1.82
caniceps	9	5	0.806 +/- 0.120	0.0010 +/- 0.0007	-0.08
dorsalis	10	6	0.889 +/- 0.075	0.0009 +/- 0.0006	-1.34
phaeonotus	76	27	0.934 +/- 0.013	0.0010 +/- 0.0006	-18.19*
fulvescens	8	3	0.679 +/- 0.122	0.0005 +/- 0.0004	0.83
insularis	23	9	0.889 +/- 0.032	0.0016 +/- 0.0009	0.04
bairdi	11	3	0.473 +/- 0.162	0.0004 +/- 0.0003	0.90
alticola	15	7	0.895 +/- 0.043	0.0012 +/- 0.0007	-0.36
vulcani	2	2	1.000 +/- 0.500	0.0004 +/- 0.0005	0.00

Genome-wide structure from principal components analysis

A PCA of the entire SNP dataset revealed a pattern of clustering remarkably congruent with both the geographic distribution and phenotypic characteristics of junco taxa (Fig. 1.3), especially when plotting the PC1 again the PC3 (Fig. 1.3A). The analysis showed a clear separation between the yellow and dark-eyed juncos, but also between the northernmost forms *aikeni* and *hyemalis* with respect the rest of the groups. Structure was evident even between the subspecific groups of the Oregon junco (*montanus* and *thurberi*) and those from the yellow-eyed junco (*palliatus* and *phaeonotus*), showing relatively less separation among groups from the middle of the geographic distribution.

Phylogenomic reconstruction

A maximum likelihood phylogeny based on 32,674 selectively neutral SNP loci corresponding to the top 25% of the global-F_{ST} distribution was consistent with the mtDNA phylogeny in showing a marked contrast between isolated southern lineages and recently diverged forms in mainland North America (Fig. 1.4A). Forms *bairdi* and *insularis* showed the longest branches, whereas Guatemalan *alticola*, which appeared as sister to *insularis* in the mtDNA phylogeny, shows a shorter branch within the mainland cluster (Fig. 1.4A). Interestingly, in this phylogeny *fulvescens* from Chiapas appears as sister to *alticola*, whereas in the mtDNA phylogeny it clusters with the clade composed of mainland yellow-eyed juncos and dark-eyed juncos. The remaining taxa show a pattern of reciprocal monophyly, although node support values were low. To increase resolution among recently diverged northern forms, we analyzed these taxa separately (excluding *bairdi*, *insularis*, *alticola* and *fulvescens*). The resulting ML unrooted tree revealed a pattern of strong reciprocal monophyly among junco taxa with bootstrap values ranging from 87 to 100%, except the nodes for *dorsalis* and *mearnsi* which received values of 63% and 62%, respectively (Fig. 1.4B). Support for nodes among the main clades was generally strong, except for that between *hyemalis* and *aikeni*, which was 30%. The dataset that included only the top 10% of SNPs from the global F_{ST} distribution showed slightly reduced phylogenetic resolution, with one *caniceps* individuals grouping with the *mearnsi* group (Fig. 1.5C). The dataset that included all SNPs presented less resolution, separating yellow-eyed juncos from dark-eyed juncos and

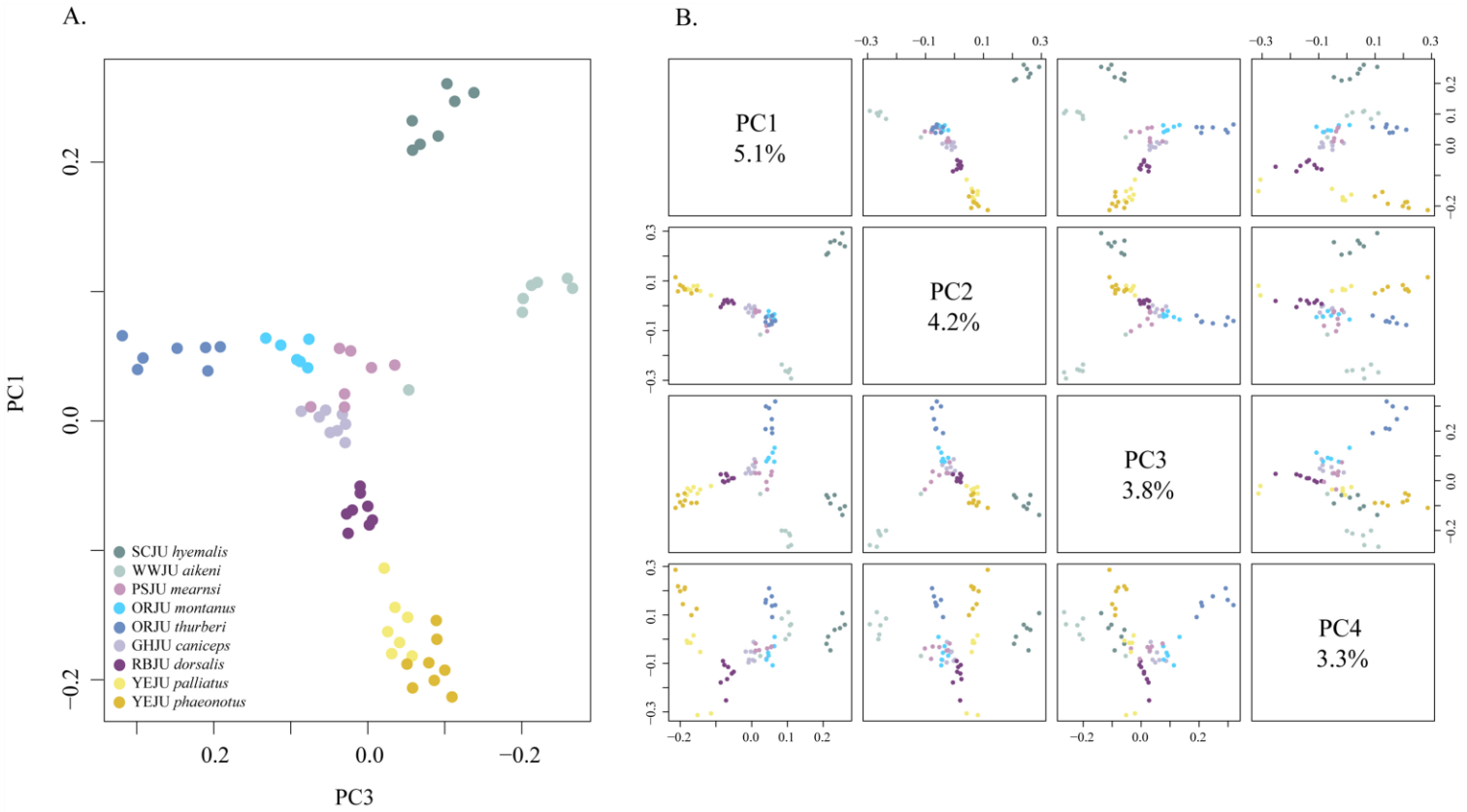


Figure 1.3. Genetic structure of northern junco forms based on a principal components analysis of selectively neutral genome-wide SNPs. (A) Plot of the PC1 against the PC3. (B) Pairwise plots of the first four principal components. Marker colors correspond to those on the range map on Fig. 1.1 of the General Introduction, except for ORJU and YEJU, which are divided into two additional taxa shown by different shades.

resolving *dorsalis* and *hyemalis* forms while showing reduced resolution for the remaining recently diverged clades (Fig. 1.5A).

Rooting the SNP tree with *fulvescens* from Chiapas revealed a striking pattern of diversification consistent with a northward sequence of cladogenetic events (Fig. 1.6). Thus, relative to *fulvescens*, *phaeonotus* (central Mexico) appears as the most basal taxon, followed by *palliatus* (northern Mexico), *dorsalis* (Arizona and New Mexico), *caniceps* (Rocky Mts.), and finally the most recently diversified taxa in the north (*oreganus*, *hyemalis* and *aikeni*) with gradually decreasing node support.

The phylogeny based on a randomized dataset (global Fst for top 25% = 0.036-0.15, Table 1.2) showed little structure and failed to produce monophyletic clusters among randomly assigned individuals (Fig. 1.7), indicating that the phylogenetic patterns recovered with the SNP datasets are robust.

Phenotypic differentiation and correlations with genetic distance

The plot of the two first discriminant functions from a DFA of morphometric variables revealed considerable differentiation among southern lineages, in contrast to northern forms, which generally showed extensive overlap (Fig. 1.8A). In contrast, colorimetric variables did show a clear separation between some northern taxa, yet widespread overlap among most of the older lineages (Fig. 1.8B), with the exception of *insularis* (GUJU), which clusters next to distantly-related *mearnsi* (PSJU), illustrating a case of evolutionary convergence in plumage coloration within the junco system as reported previously (Aleixandre et al. 2013). Heat-map plots illustrate the markedly different correlation between genetic distances and morphometric and colorimetric distances, respectively (Fig. 1.8C, D). The pattern is particularly striking for northern forms (upper left quadrant of the matrices), which show little genetic divergence (yellow cells below the diagonal) and morphological divergence (predominantly yellow cells above the diagonal in Fig. 1.8C), and marked plumage color divergence (predominantly orange and red cells above the diagonal in Fig. 1.8D). A Mantel test revealed a significant correlation between matrices of genetic and morphological distances ($r = 0.595$, $P < 0.001$), but not between genetic and colorimetric distances ($r = -0.169$, $P = 0.837$).

Figure 1.4. Phylogenetic relationships among junco forms based on genome-wide SNP markers. Unrooted maximum likelihood phylogenies of (A) all junco taxa based on 32,674 selectively neutral SNP loci, and (B) recently diverged northern junco forms only (excluding highly divergent taxa) based on 25,286 selectively neutral SNP loci. See Methods section for SNP filtering criteria. Branch colors correspond to those on Fig. 1.1.

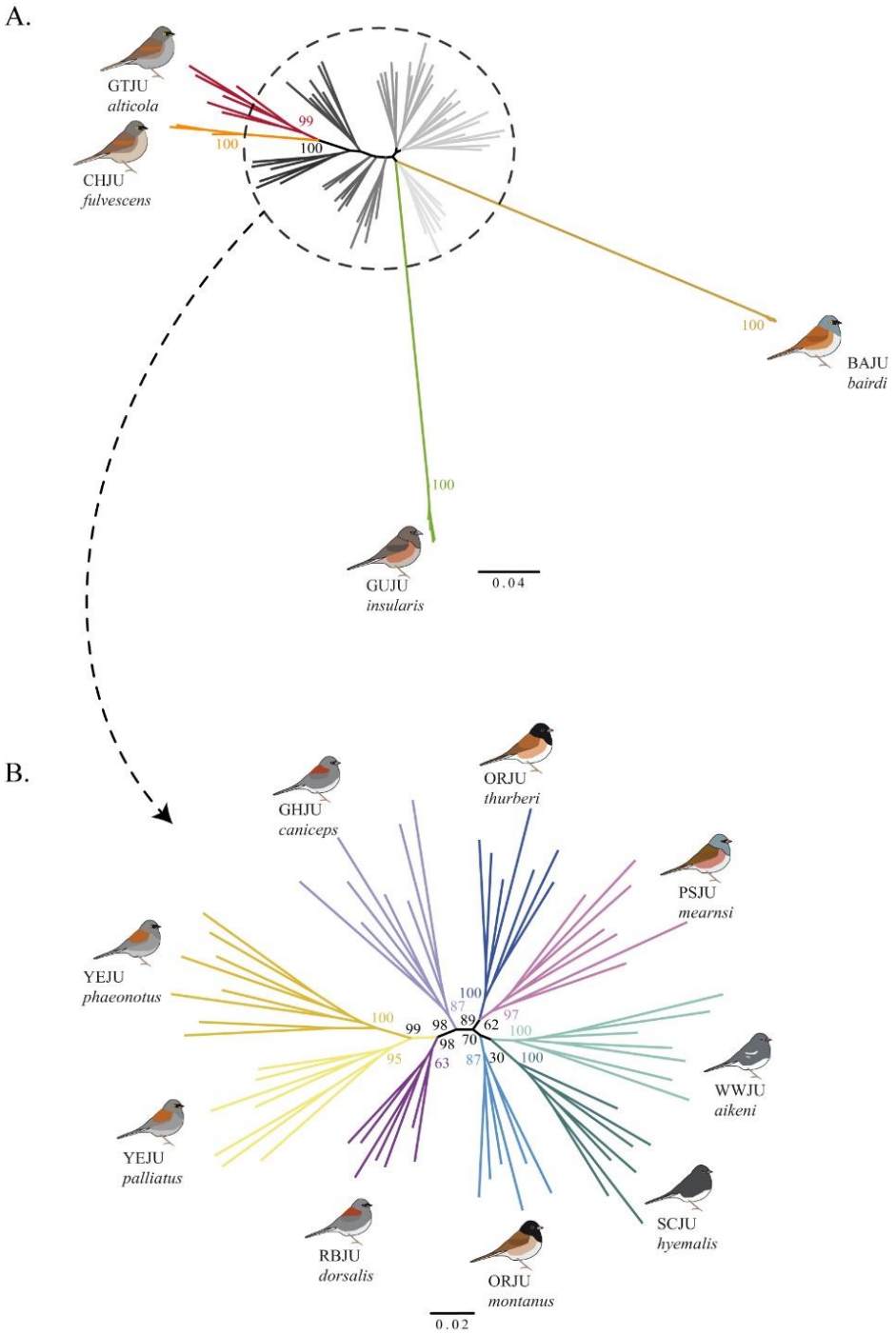


Figure 1.5. Variation in phylogenetic signal among SNP datasets from across the global-F_{ST} distribution. From an initial dataset of 101,145 neutral SNPs, shown are unrooted maximum likelihood phylogenies of northern junco forms (right panels) based on SNP matrices of different ranges within the global-F_{ST} distribution (left panels), corresponding to 100% (A), top 25% (B), and top10% SNPs from the global-F_{ST} distribution (C). Blue bars correspond to the SNPs used, gray bars correspond to the SNPs excluded (if any), and black lines correspond to the accumulated percentage of SNPs over F_{ST} classes.

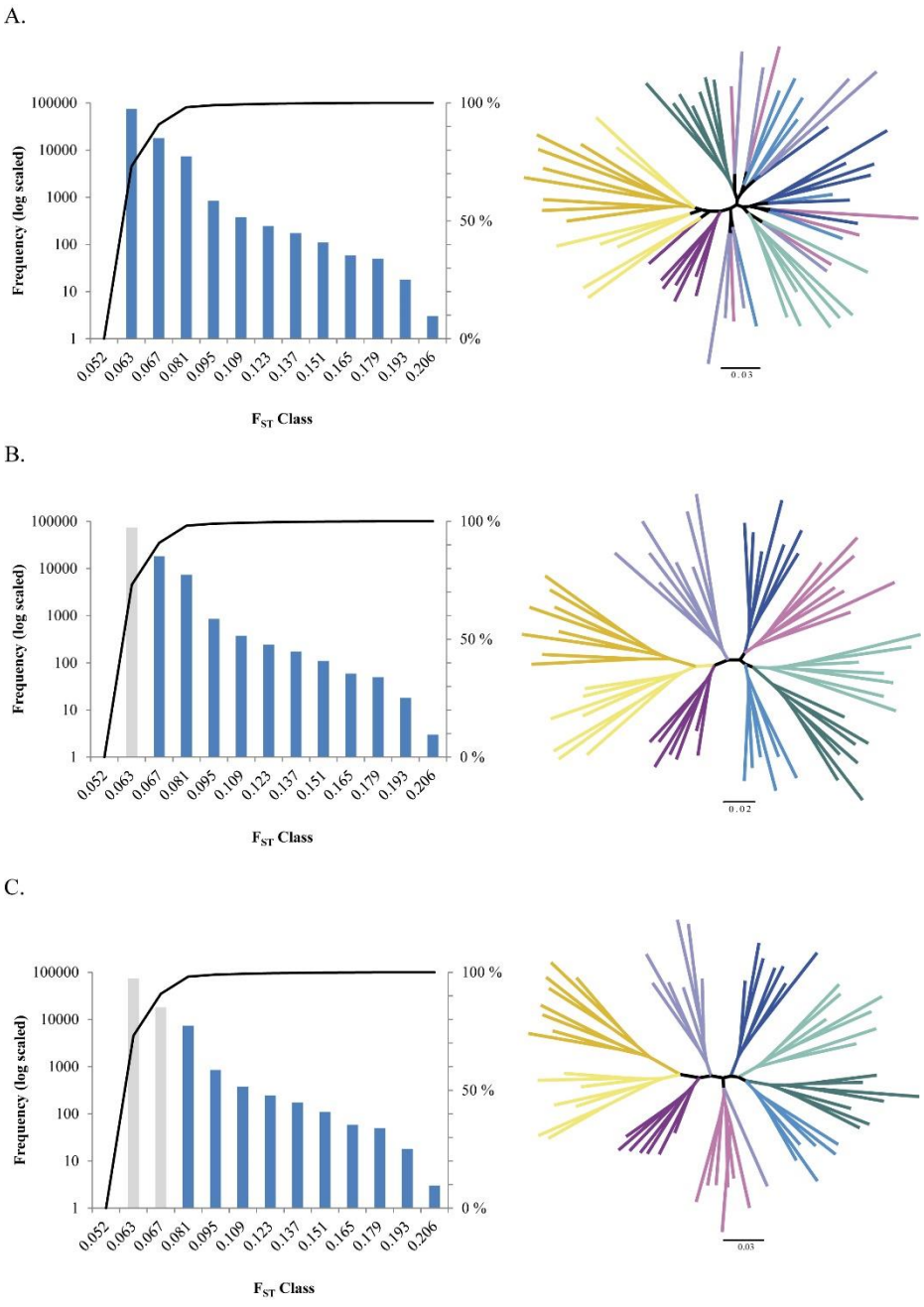
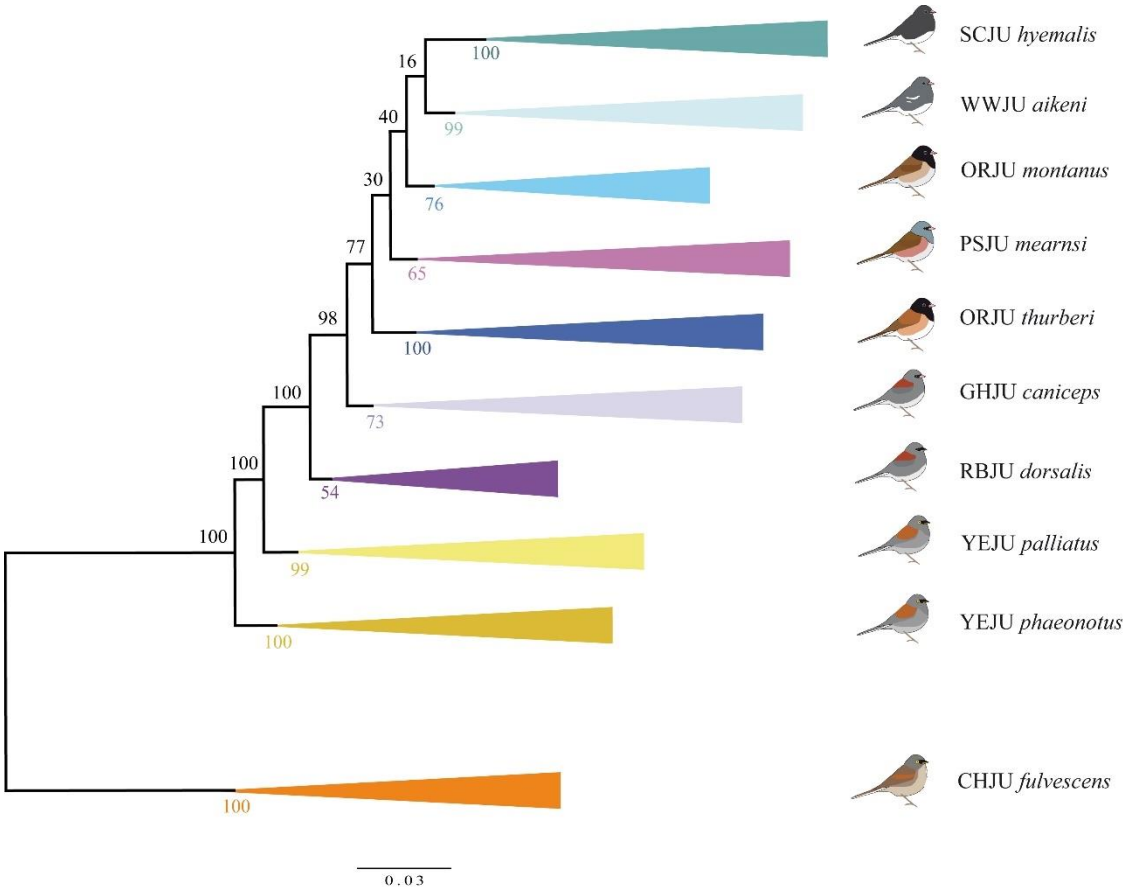


Figure 1.6. Phylogenetic footprint of a recent northward expansion using SNP data. Maximum likelihood phylogeny of northern junco forms based on a subset of neutral SNP loci restricted to the 25% with highest global F_{ST} (26,393 loci, F_{ST} = 0.09 to 0.26). The Chiapas junco (*fulvescens*, CHJU) from southern Mexico is used to root the tree.



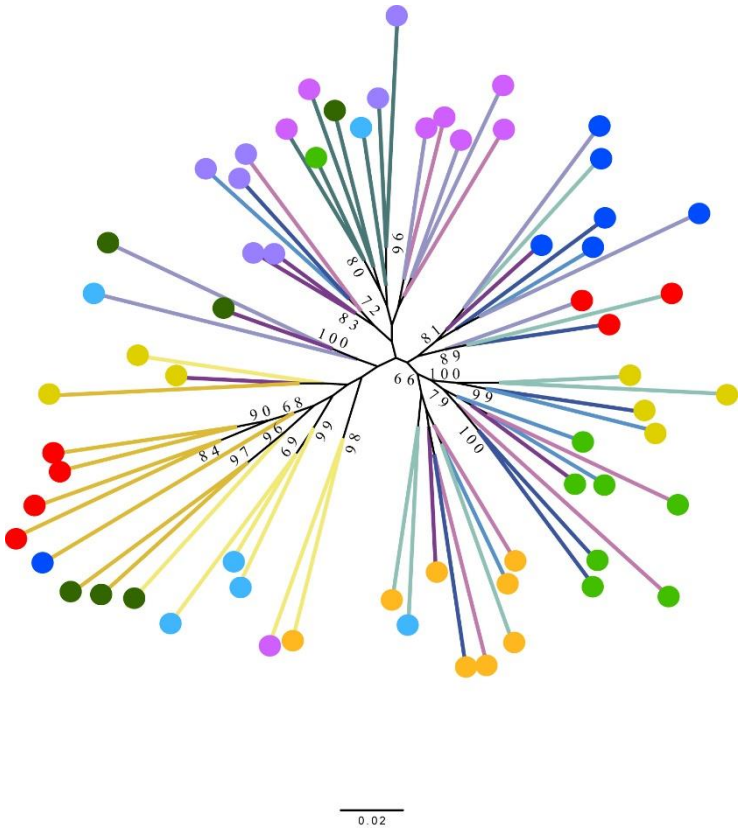
Discussion

Young and old junco clades: rapid diversification vs. long-term stasis

Our results reveal the existence of closely related evolutionary lineages within the genus *Junco* that have diversified at strikingly different rates. Lineages that have been geographically isolated in oceanic or sky islands for over two hundred thousand years have diverged markedly from each other in neutral genetic markers, but relatively little in plumage color and pattern. In sharp contrast, the postglacial colonization of North America by one of the yellow-eyed lineages from southern

Mexico, has caused a burst of diversification that has resulted in no less than six phenotypically and genetically distinct dark-eyed junco taxa. These young lineages differ most prominently in plumage coloration and patterning, both between each other and with respect to ancestral yellow-eyed taxa to the south, yet show relatively limited divergence in morphological traits. Morphological divergence in

Figure 1.7. Maximum likelihood phylogeny based on a randomized dataset. To test the significance of the phylogenetic signal in the Fig. 1.4B phylogeny, individuals were randomly assigned to hypothetical populations of similar sample size, and were analyzed following the same protocol [excluding outliers with BayeScan (Foll and Gaggiotti 2008) and using only the top-25% F_{ST} SNPs among groups]. Colored dots at branch tips represent random populations, and branch colors correspond to the real junco forms as seen in Figure 1.1 from the General Introduction. The figure shows that phylogenetic signal is low, and colored dots do not form monophyletic groups. In fact, some groups are robust to the randomization of individuals, like *hyemalis* (dark-green branches), or yellow-eyed junco forms *palliatus* and *phaeonotus* (yellow and ochre branches) which cluster together despite being assigned to different populations.



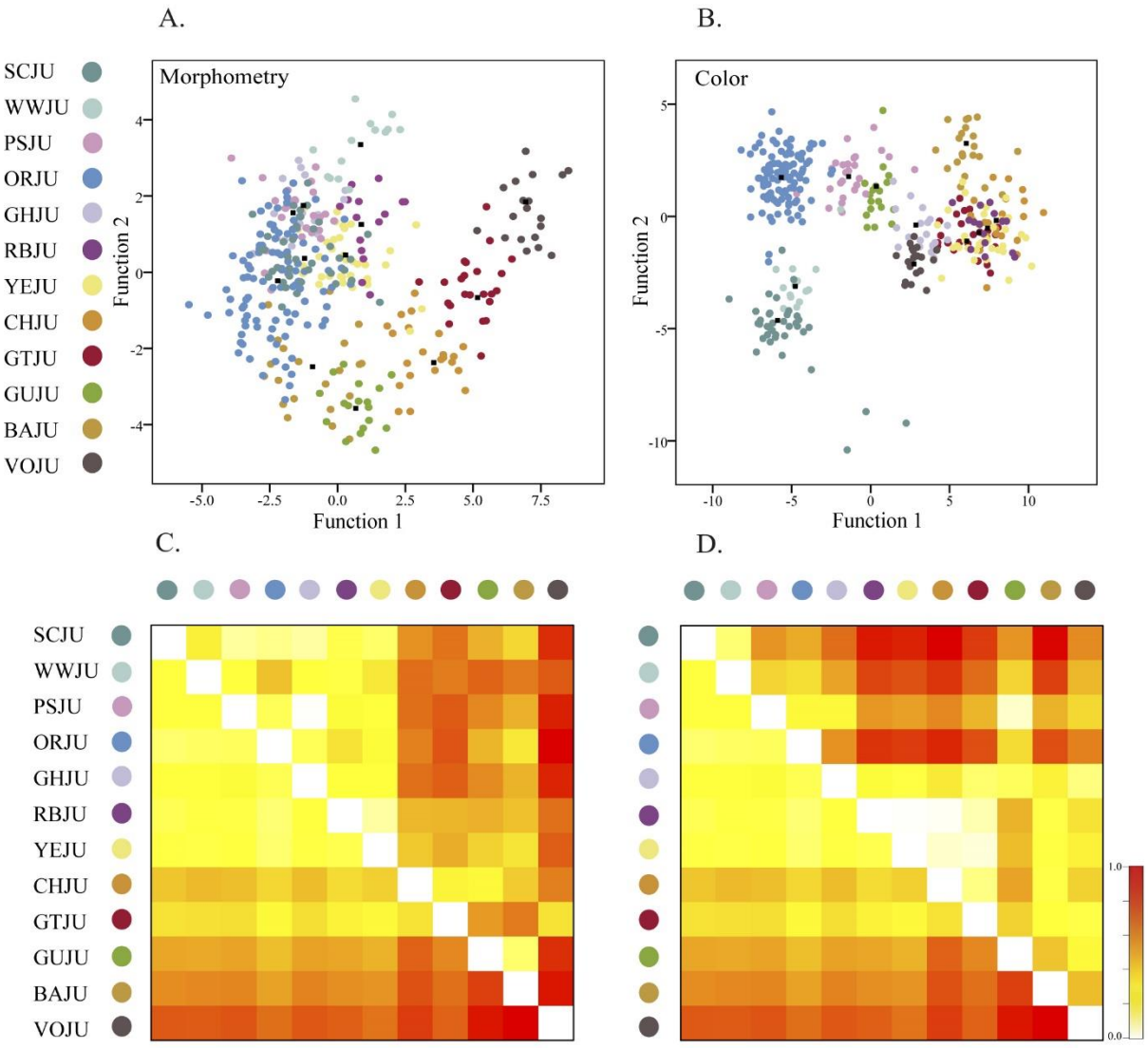
older lineages is marked but variable, with *bairdi* showing average characters compared to the significantly enlarged *alticola*, which inhabits the highlands of Guatemala and eastern Chiapas at nearly 4,000 m elevation, and the small size and elongated bill of *insularis* on Guadalupe Island, a species that has likely been subjected to strong divergent selection to feed on a limited set of insular food types (Aleixandre et al. 2013). Overall, it is apparent that within juncos, species richness is clearly decoupled from clade age, and phenotypic diversification appears to show a better correlation with the rate of cladogenesis than with lineage age, patterns which have been shown to emerge at macroevolutionary scales (Rabosky and Adams 2012; Rabosky et al. 2012; Rabosky et al. 2013).

Understanding the micro-evolutionary mechanisms that give rise to such different rates of diversification within a single bird genus require careful testing of the relative roles of natural and sexual selection, and even drift, in driving divergence (see Chapters II and III), yet some prominent differences between northern and southern lineages seem relevant. One is that old southern lineages have remained isolated in areas with relatively unchanging habitat in climatically stable regions over many thousands of years, in contrast to young northern lineages, which have diversified across a large continental land mass spanning a latitudinal gradient with marked environmental variation. In both cases, the patchy distribution of suitable junco habitat in many parts of the junco range, where populations are isolated in high-elevation sky islands, may have promoted periods of complete isolation, where both selection and drift can act quickly to drive phenotypic divergence. In any case, our results are consistent with studies showing that the evolvability of clades may be related to their capacity to diversify across climatic gradients, as shown for example in the tanagers (Thraupidae), where species richness is correlated with climatic niche evolution (Title and Burns 2015), or in African lake cichlids, where a combination of ecological factors and sexual selection explain the probability of adaptive radiation in African lakes (Wagner et al. 2012).

Rapid postglacial speciation along a latitudinal environmental gradient

Our mtDNA and genome-wide datasets confirm a striking case of rapid phenotypic diversification in dark-eyed juncos from a yellow-eyed junco ancestor.

Figure 1.8. Phenotypic differences among junco forms and their correlation with genetic distance. The first two discriminant functions in a discriminant function analysis (DFA) based on morphological variables (A) and plumage color variables (B) show patterns of phenotypic differentiation among junco morphs. Heat-map plots showing the correlation between pairwise SNP-based genetic distances among forms (below diagonal) and Euclidean distances among DFA centroids (above diagonal) for morphometric (C) and colorimetric (D) variables. Morphometric and colorimetric divergence values between forms show generally opposite patterns among old and recently diverged junco lineages.



Phylogeographic data based on several mtDNA coding regions reveal a clear population expansion of dark-eyed junco morphs across North America as evidenced by star-like haplotype phylogenies and significant tests of demographic expansion, confirming previous results obtained with smaller datasets (Milá et al. 2007). Numerous studies on avian taxa have documented similar postglacial expansions using phylogeographic and genetic diversity data (Seutin et al. 1995; Milá et al. 2000; Milá et al. 2006; Hansson et al. 2008; Malpica and Ornelas 2014; Alvarez et al. 2015), although none has shown levels of phenotypic diversification comparable to those in juncos. Our coalescence analyses provide estimates of the time since the junco expansion ranging from 41,300 to 121,800 years ago when measured at the *fulvescens*-northern juncos node, and from 6,800 to 88,600 years ago when computing the TMRCA for the northern taxa. However, given the date of Last Glacial Maximum 18,000 years ago, the IMa2 mean value estimate of 16,132 years ago is clearly more consistent with available geo-climatic data than the 58,300 years estimated with BEAST, which corresponds to a period when suitable junco habitat was lacking in most of the North American continent.

In addition to mtDNA sequence data, our analysis of genome-wide SNP loci provides unprecedented insight into the rapid junco radiation. As in previous studies using genome-wide SNPs to reconstruct the evolutionary history of recently radiated complexes, we obtained a better resolution of phylogenetic relationships among phenotypically differentiated taxa when using loci in the top 25% of a global-F_{ST} distribution than when including all available loci. Recent studies by Puebla *et al.* (2014) and Keller *et al.* (2013) obtained the best phylogenetic signal among recently diverged fish taxa when using only the top 10% and 25% of the global- F_{ST} distribution, respectively. The lack of complete congruence between high-F_{ST} loci and the entire dataset is likely due to the stochastic nature of the coalescence process among loci, with the probability of finding loci with histories that are incongruent with the species tree increasing when large numbers of loci are used. Using high- F_{ST} loci among genetically and phenotypically differentiated groups (following the exclusion of outlier loci potentially under selection), is thus a way of reducing the influence of loci with incongruent histories due to incomplete lineage sorting, drift, divergent selection or balancing selection.

The resolution achieved in our PCA and phylogenomic analyses clarifies the evolutionary history of the juncos in three major ways. First, it confirms the young age of dark-eyed junco lineages relative to southern lineages, showing short-branched but reciprocally monophyletic clades for the different forms. This demonstrates that continent-wide sharing of mtDNA haplotypes in North America represents shared ancestral polymorphism due to recent divergence, and thus rules out the alternative explanation of introgressive hybridization and mtDNA capture in the early stages of the northward expansion, as has been shown to have occurred in other species with similar distributional histories (Milá et al. 2011). In fact, our SNP data reveal a possible case of past mtDNA introgressive hybridization in *fulvescens* of Chiapas, in southern Mexico. Individuals of *fulvescens* carry mtDNA haplotypes closely related to North American forms, yet genome-wide SNP data shows they are actually more closely related to *alticola* of Guatemala, which is more consistent with geography and phenotype. A likely scenario causing this example of mito-nuclear discordance is the introduction of mtDNA haplotypes into the *fulvescens* population through incidental hybridization with few *phaeonotus* females, followed by the subsequent fixation of those haplotypes in drift-prone small highland populations.

Second, the dark-eyed junco phylogeny based on neutral genome-wide SNPs recovers with remarkable precision the south-to-north sequence of cladogenetic events, from yellow-eyed forms in Mexico (*phaeonotus* and *palliatus*), through the first dark-eyed forms in the southern USA (*dorsalis* and *caniceps*), to the most recently evolved forms in the north (*hyemalis* and *oreganus*), once again confirming the role of a rapid postglacial expansion in driving the fast radiation. Consistent with this scenario are the node support values on the tree, which decrease gradually with increasing latitude, showing the lowest support for the most recently evolved northern forms.

And third, the clear pattern of genetic structure and even reciprocal monophyly among young dark-eyed junco clades indicates that they represent relatively homogeneous genetic entities with limited gene flow among them. This degree of differentiation among recently diverged forms is particularly striking given that the

SNPs sampled correspond to only about 0.4% of the junco's ~1.2 Gb genome, and that outlier loci potentially under selection were filtered out for this analysis. Unlike other avian models in which divergence appears to be limited to a reduced number of regions or "genomic islands of divergence" (Ellegren et al. 2012; Poelstra et al. 2014) the clear separation of taxa using a small subset of neutral genome-wide markers suggests that divergence has taken place at the level of the entire genome. This may indicate a role for local adaptation of junco forms to environmental conditions along the latitudinal gradient, which is thought to result from relatively weak multifarious selection acting on many loci across the genome as opposed to strong selection on a few loci (Dambroski and Feder 2007; Michel et al. 2010; Parchman et al. 2013; Egea-Serrano et al. 2015). When selection acts on many loci, the combined effect of divergent selection can drive a global reduction in effective gene flow, thus favoring the establishment of new mutations and leading to "genome hitchhiking" (Feder and Nosil 2010; Feder et al. 2013), a process with the potential to drive genome-wide divergence even in the early stages of population divergence (Michel et al. 2010; Roesti et al. 2012). Thus, a relatively small number of markers may be sufficient to detect genome-wide divergence caused by multifarious selection driven by local adaptation. Alternatively, the low-level, genome-wide divergence found among junco forms could have been caused by drift if the original populations suffered strong reductions in effective population size during geographic isolation. The large number of differences encountered across the genome within a few thousand years, and the relatively slow substitution rate under drift compared to selection (even low-level multifarious selection), suggests drift may be a less parsimonious explanation for the observed patterns of variation, but higher-density genome scans will be necessary to further explore the genomic landscape of this unique radiation in order to assess the relative importance of drift, multifarious selection, and strong selection acting on specific loci across the junco genome.

The junco radiation stands out as one of the fastest known in vertebrates. Other famous avian radiations like those of *Zosterops* white-eyes (Moyle et al. 2009), *Setophaga* warblers (Lovette et al. 2010), or Hawaiian depranidid honeycreepers (Lerner et al. 2011), gave rise to vast number of species, but over relatively longer

periods spanning several million years. One case that is remarkably comparable in timescale and number of resulting taxa is that of southern capuchinos, a clade of South American *Sporophila* seedeaters, thought to have originated within the last 40,000 years and whose eight members differ conspicuously in melanic plumage coloration (Campagna et al. 2015). Unlike juncos, however, most morphs are largely sympatric across ecologically similar habitat, and signal for genetic structure using similar genome-wide markers to the ones used here, is very weak (Campagna et al. 2015). This lack of phylogenetic signal in random genome-wide SNPs points to a major role for a few loci of strong effect in determining southern capuchino phenotypes, with the rest of the genome showing no differentiation due to either shared ancestral polymorphism or widespread gene flow in unselected genomic regions. Future research into the evolutionary mechanisms involved and the respective genomic signatures in both systems, will help us understand the relative roles of gene flow, selection and historical context in driving fast divergence and speciation (see next chapters).

A potential role for sexual selection

The marked differences in plumage pattern and color among recently diversified dark-eyed junco morphs suggest the role of sexual selection, an important mechanism in divergence and speciation (Lande 1981; West-Eberhard 1983; Wilson et al. 2000; Coyne and Orr 2004; Ritchie 2007). Evidence for the direct role of sexual selection in modifying junco phenotypes is limited, but studies at the population level have shown that sexual selection can modify the amount of white on tail feathers in slate-colored and Oregon juncos, a trait known to be involved in mate choice (Hill et al. 1999; Yeh 2004; McGlothlin et al. 2005; Price 2008). At the level of the entire complex, a testable prediction of the action of sexual selection in evolutionary divergence is the appearance of sexual dimorphism in secondary sexual characters (Darwin 1872). Based on the examination of thousands of junco specimens, Miller (1941) assessed the amount of sexual dimorphism in head color among the different junco taxa, and he concluded that dimorphism appears to increase with latitude: it is absent in *vulcani*, *bairdi*, *alticola*, and *fulvescens*, very low in *phaeonotus, palliatus, dorsalis,* and *insularis*, slight in *caniceps*, and clearly present in the various *oreganus* and *hyemalis* taxa. Although it remains to be properly

quantified, this pattern would be congruent with both latitude and the phylogenetic pattern we have reported, suggesting that sexual selection may have played a role in the fast diversification of northern forms. Indeed, an interaction between sexual selection and local adaptation has been shown to promote speciation in recent models, showing that sexual ornaments can function as indicators of local adaptation and increase disruptive selection driven by ecological factors (van Doorn et al. 2009). We test this hypothesis in detail in Chapter II.

Taxonomic implications

Diagnosability through plumage coloration has been and remains the single most important criterion in avian species designation (Watson 2005; Sangster et al. 2011). This is the case even though phenotypic divergence has often been shown to be a poor predictor of evolutionary history (Olsson et al. 2005; Navarro-Sigüenza et al. 2008; Gill 2014). Juncos are a prime example, and our results reveal major incongruences between the group's evolutionary history and current taxonomy. Traditional taxonomic classifications based on eye color (yellow vs. dark) are clearly paraphyletic, and counterintuitively, the more uniform-looking yellow-eyed juncos in the south turned out to be the most evolutionarily divergent lineages, with the most distinct forms among dark-eyed juncos of North America being the youngest. Even before these findings, junco taxonomy has never been simple, and the lack of consensus on how junco diversity must be classified has led to a convoluted taxonomic history (Ketterson and Atwell 2016; Milá et al. 2016).

Based on the combination of genetic, genomic and phenotypic data presented herein, we propose that Baird's junco (*J. bairdi*) and the Guatemala junco (*J. alticola*) be recognized as separate species, with the latter including the Chiapas junco (subspecies *J. a. fulvescens*). The island junco (*Junco insularis*) from Guadalupe Island was already proposed as a separate species based on a previous study (Aleixandre et al. 2013), and it has been recently adopted by the American Ornithologists' Union (Chesser et al. 2014) and the IOC World Bird List (Gill and Donsker 2016). Genetic and phenotypic data also provide support for maintaining the yellow-eyed junco (*J. phaeonotus*) as a separate species that includes the two described subspecies *J. p. phaeonotus* and *J. p. palliatus*. With respect to the recently diverged dark-eyed junco

forms that breed in the USA and Canada, currently grouped into a single species (*Junco hyemalis*), we believe that a thorough reexamination of geographic variation in phenotypic traits in light of the genomic data presented here will result in the recognition of at least four species-level taxa, namely the gray-headed junco (*J. caniceps*), the Oregon junco (*J. oreganus*), the white-winged junco (*J. aikeni*) and the slate-colored junco (*J. hyemalis*). This taxonomic arrangement would in fact be quite similar to the one proposed by Alden Miller (1941), whose keen observations and tireless efforts may well be vindicated now, over seventy years later, by the genomic data he lacked. However, given the complex patterns of phenotypic variation at different geographic scales and the varying degrees of intergradation among some forms within the group, issuing conclusive and detailed taxonomic recommendations on this group will require in-depth descriptions, analyses and discussions that are beyond the scope of the present paper, and thus will be presented elsewhere.

Conclusion

Our analyses confirm a striking case of rapid postglacial speciation in the genus *Junco* and reveal marked differences in the rate and degree of phenotypic diversification among clades. The rapid diversification of distinctly colored dark-eyed junco forms across North America took place since the Last Glacial Maximum just 18,000 years ago. In contrast, isolated lineages in the south show relative phenotypic stasis, with plumage color showing little change over hundreds of thousands of years. The pattern of genome-wide divergence among northern forms despite their recent divergence, suggests that in addition to color differences, which are known to be controlled by a small number of pigmentation loci, junco forms have diverged in many other traits likely related to adaptation to local environmental conditions. In-depth genome scans will be necessary to identify genomic regions under selection and reveal the genomic architecture of rapid diversification. Our within-genus study reveals a compelling example of speciation-in-action in juncos, where the combined effect of historical and selective factors has produced one of the fastest cases of speciation known in vertebrates.

References

Aleixandre, P., J. Hernández-Montoya, and B. Milá. 2013. Speciation on oceanic islands: rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: *Junco*). PLOS One 8(5):e63242.

Alvarez, S., J. F. Salter, J. E. McCormack, and B. Milá. 2015. Speciation in mountain refugia: phylogeography and demographic history of the pine siskin and black-capped siskin complex. Journal of Avian Biology 47:335-345.

Baele, G., P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, and A. V. Alekseyenko. 2012. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. Molecular biology and evolution 29:2157-2167.

Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution 16:37-48.

Campagna, L., I. Gronau, L. F. Silveira, A. Siepel, and I. J. Lovette. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular ecology 24:4238-4251.

Chesser, R. T., R. C. Banks, C. Cicero, J. L. Dunn, A. W. Kratter, I. J. Lovette, S. Navarro, xfc, A. G. enza, P. C. Rasmussen, J. V. Remsen, J. D. Rising, D. F. Stotz, and K. Winker. 2014. Fifty-fifth supplement to the American Ornithological Society's Check-list of North American Birds. The Auk 131:i-xv.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Currie, D. J., G. G. Mittelbach, H. V. Cornell, R. Field, J. F. Guégan, B. A. Hawkins, D. M. Kaufman, J. T. Kerr, T. Oberdorff, and E. O'Brien. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecology letters 7:1121-1134.

Dambroski, H. R. and J. L. Feder. 2007. Host plant and latitude-related diapause variation in *Rhagoletis pomonella*: a test for multifaceted life history adaptation on different stages of diapause development. Journal of Evolutionary Biology 20:2101-2112.

Darwin, C. 1872. The Origin of Species by means of Natural Selection, or the preservation of favoured Races in the struggle for Life. Sixth edition, with additions and corrections.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular biology and evolution 22:1185-1192.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969-1973.

Eberhard, J. R. and E. Bermingham. 2005. Phylogeny and comparative biogeography of *Pionopsitta* parrots and *Pteroglossus* toucans. Molecular Phylogenetics and Evolution 36:288-304.

Egea-Serrano, A., S. Hangartner, A. Laurila, and K. Räsänen. 2015. Multifarious selection through environmental change: acidity and predator-mediated adaptive divergence in the moor frog (*Rana arvalis*). Proceedings of the Royal Society B 281:20133266.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756-760.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics 1:47-50.

Feder, J. L., S. M. Flaxman, S. P. Egan, A. A. Comeault, and P. Nosil. 2013. Geographic mode of speciation and genomic divergence. Annual Review of Ecology, Evolution, and Systematics 44:73-97.

Feder, J. L. and P. Nosil. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729-1747.

Foll, M. 2012. Bayescan v2. 1 user manual. Ecology 20:1450-1462.

Foll, M. and O. Gaggiotti. 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180:977-993.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Gill, F. and D. Donsker. 2016. IOC World Bird List (v 6.2). Preneseno iz: http://www. worldbirdnames. org/, pridobljeno 18:2016.

Gill, F. B. 2014. Species taxonomy of birds: Which null hypothesis? The Auk 131:150-161.

Hackett, S. J. 1996. Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). Molecular Phylogenetics and Evolution 5:368-382.

Hansson, B., D. Hasselquist, M. Tarka, P. Zehtindjiev, and S. Bensch. 2008. Postglacial colonisation patterns and the role of isolation and expansion in driving diversification in a passerine bird. PLoS One 3:e2794.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. PLoS Biology 2:1657-1663.

Hewitt, G. M. 1996. Some genetic consequences of ice ages and their role in divergence and speciation. Biological Journal of the Linnean Society 58:247-276.

Hey, J. 2010. Isolation with migration models for more than two populations. Molecular Biology and Evolution 27:905-920.

Hill, J. A., D. A. Enstrom, E. D. Ketterson, V. J. Nolan, and C. Ziegenfus. 1999. Mate choice based on static versus dynamic secondary sexual traits in the dark-eyed junco. Behavioral Ecology 10:91-96.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Keller, I., C. E. Wagner, L. Greuter, S. Mwaiko, O. M. Selz, A. Sivasundar, S. Wittwer, and O. Seehausen. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular ecology 22:2848-2863.

Ketterson, E. D. and J. W. Atwell. 2016. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco. University of Chicago Press.

Kimball, R. T., E. L. Braun, F. K. Barker, R. C. Bowie, M. J. Braun, J. L. Chojnowski, S. J. Hackett, K.-L. Han, J. Harshman, and V. Heimer-Torres. 2009. A well-tested set of primers to amplify regions spread across the avian genome. Molecular Phylogenetics and Evolution 50:654-660.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the USA 78:3721-3725.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular biology and evolution 29:1695-1701.

Lerner, H. R. L., M. Meyer, H. F. James, M. Hofreiter, and R. C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers. Current Biology 21:1838-1844.

Lischer, H. E. L. and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Lovette, I. J., J. L. Pérez-Emán, J. P. Sullivan, R. C. Banks, I. Fiorentino, S. Córdoba-Córdoba, M. Echeverry-Galvis, F. K. Barker, K. J. Burns, J. Klicka, S. M. Lanyon, and E. Bermingham. 2010. A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). Molecular Phylogenetics and Evolution 57:753-770.

Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. Evolution 64:2731-2745.

Malpica, A. and J. F. Ornelas. 2014. Postglacial northward expansion and genetic differentiation between migratory and sedentary populations of the broad-tailed hummingbird (*Selasphorus platycercus*). Molecular Ecology 23:435-452.

McGlothlin, J. W., P. G. Parker, V. J. Nolan, and E. D. Ketterson. 2005. Correlational selection leads to genetic integration of body size and an attractive plumage trait in dark-eyed juncos. Evolution 59:658-671.

Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. Proceedings of the National Academy of Sciences 107:9724-9729.

Milá, B., P. Aleixandre, S. Alvarez-Nordström, J. McCormack, E. Ketterson, and J. Atwell. 2016. More than meets the eye: Lineage diversity and evolutionary history of Dark-eyed and Yellow-eyed juncos. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco (ED Ketterson and JW Atwell, Editors). University of Chicago Press, Chicago, Illinois, USA:179-198.

Milá, B., D. J. Girman, M. Kimura, and T. B. Smith. 2000. Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proceedings of the Royal Society Biological Sciences Series B 267:1033-1040.

Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society B-Biological Sciences 274:2653-2660.

Milá, B., T. B. Smith, and R. K. Wayne. 2006. Postglacial population expansion drives the evolution of long-distance migration in a songbird. Evolution 60:2403-2409.

Milá, B., D. P. L. Toews, T. B. Smith, R. K. Wayne, J. Lake, and N. F. Dixie. 2011. A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (Aves : *Dendroica coronata*). Biological Journal of the Linnean Society 103:696-706.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1-8. Gateway Computing Environments Workshop (GCE), 2010. Ieee.

Moyle, R. G., C. E. Filardi, J. J. Smith, and J. M. Diamond. 2009. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". Proceedings of the National Academy of Sciences of the United States of America 106:1863-1868.

Navarro-Sigüenza, A. G., A. T. Peterson, A. Nyari, G. M. Garcia-Deras, and J. Garcia-Moreno. 2008. Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. Molecular Phylogenetics and Evolution 47:21-35.

Nei, M. 1987. Genetic distance and molecular phylogeny. Pp. 193-224 *in* N. Ryman, and F. Utter, eds. Population genetics and fisheries management. Univ. of Washington Press, Seattle.

Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus, S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Oksanen, J., R. Kindt, and B. O'Hara. 2005. Vegan: R functions for vegetation ecologists. URL: http://cc.oulu.fi/jarioksa/softhelp/vegan.html.

Olsson, U., P. Alström, P. G. Ericson, and P. Sundberg. 2005. Non-monophyletic taxa and cryptic species—evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). Molecular phylogenetics and evolution 36:261-276.

Parchman, T., Z. Gompert, M. Braun, R. Brumfield, D. McDonald, J. Uy, G. Zhang, E. Jarvis, B. Schlinger, and C. Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology 22:3304-3317.

Phillimore, A. B., C. D. L. Orme, R. G. Davies, J. D. Hadfield, W. J. Reed, K. J. Gaston, R. P. Freckleton, and I. P. Owens. 2007. Biogeographical basis of recent phenotypic divergence among birds: a global study of subspecies richness. Evolution 61:942-957.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution 25:1253-1256.

Price, T. 2008. Speciation in birds. Roberts and Company, Greenwood Village, Colorado.

Puebla, O., E. Bermingham, and W. O. McMillan. 2014. Genomic atolls of differentiation in coral reef fishes (Hypoplectrus spp., Serranidae). Molecular ecology 23:5291-5303.

R_Core_Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D., F. Santini, J. Eastman, S. Smith, B. Sidlauskas, J. Chang, and M. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature Communications 4:1958.

Rabosky, D. L. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecology letters 12:735-743.

Rabosky, D. L. and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. Evolution 66:1807-1818.

Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade Age and Species Richness Are Decoupled Across the Eukaryotic Tree of Life. PLoS Biol 10:e1001381.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1. 6. Computer program and documentation distributed by the author, website http://beast.bio.ed.ac.uk/Tracer [accessed 27 July 2014].

Ricklefs, R. E. 2006. Global variation in the diversification rate of passerine birds. Ecology 87:2468-2478.

Ritchie, M. G. 2007. Sexual selection and speciation. Pp. 79-102. Annual Review of Ecology, Evolution, and Systematics.

Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Molecular Ecology 21:2852-2862.

Romanov, M. N., J. B. Dodgson, R. A. Gonser, and E. M. Tuttle. 2011. Comparative BAC-based mapping in the white- throated sparrow, a novel behavioral genomics model, using interspecies overgo hybridization. BMC Research Notes 4:12pp.-12pp.

Sangster, G., J. COLLINSON, P. A. CROCHET, A. G. Knox, D. T. Parkin, L. Svensson, and S. C. Votier. 2011. Taxonomic recommendations for British birds: seventh report. Ibis 153:883-892.

Seutin, G., L. M. Ratcliffe, and P. T. Boag. 1995. Mitochondrial DNA homogeneity in the phenotypically diverse redpoll finch complex (Aves: Carduelinae: *Carduelis flammea-hornemanni*). Evolution 49:962-973.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12:105-114.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57:758-771.

Tarr, C. L. 1995. Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. Molecular Ecology 4:527-529.

Title, P. O. and K. J. Burns. 2015. Rates of climatic niche evolution are correlated with species richness in a large and ecologically diverse radiation of songbirds. Ecology letters 18:433-440.

van Doorn, G. S., P. Edelaar, and F. J. Weissing. 2009. On the origin of species by natural and sexual selection. Science 326:1704-1707.

Wagner, C., L. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366-369.

Watson, D. M. 2005. Diagnosable versus distinct: evaluating species limits in birds. AIBS Bulletin 55:60-68.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. The Quarterly Review of Biology 58:155-183.

Wilson, A. B., K. Noack-Kunnmann, and A. Meyer. 2000. Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. Proceedings of the Royal Society of London B 267:2133-2141.

Yeh, P. J. 2004. Rapid evolution of a sexually selected trait following population establishment in a novel habitat. Evolution 58:166-174.

Yoder, J., E. Clancey, S. Des Roches, J. Eastman, L. Gentry, W. Godsoe, T. Hagey, D. Jochimsen, B. Oswald, and J. Robertson. 2010. Ecological opportunity and the origin of adaptive radiations. Journal of evolutionary biology 23:1581-1596.

Zheng, X. 2012. SNPRelate: parrallel computing toolset for genome-wide association studies. R package version 95.

CHAPTER II: A role for sexual selection in driving phenotypic differentiation in the rapid postglacial radiation of the dark-eyed junco

Museum specimens of *Junco h. dorsalis*, American Museum of Natural History, picture by Borja Milá

Guillermo Friis, Borja Milá

Abstract

High variability in secondary sexual traits is a common pattern among species in recent evolutionary radiations, suggesting that sexual selection may be especially important in the early stages of speciation. However, the relative role of sexual selection in promoting population differentiation and its relation with ecological factors remains controversial and difficult to assess in natural systems. Comparative analyses of sexual dichromatism —a widely used proxy of the intensity of sexual selection— with patterns of neutral and ecological differentiation across lineages of different age and degree of phenotypic diversity may help to understand the role of sexual selection in early divergence and speciation. The genus *Junco* (Aves: Emberizidae) include up to five highly divergent, independent lineages from Central America and no less than six recently diversified lineages ranging from northern Mexico through Canada as the group recolonized North America since the last glacial maximum (c.a. 18,000 years ago). The *Junco* complex shows remarkable diversity in plumage coloration and patterning, but also major habitat differences across its distribution. Here we use genomewide single nucleotide polymorphisms (SNPs) along with spectral and morphological data to investigate the roles of natural and sexual selection across *Junco* lineages, and infer how they may have contributed to their radiation. Our results show an accelerated rate of divergence in sexually selected characters relative to ecologically relevant morphological traits. Using a synthetic index of sexual dichromatism comparable across lineages, we found a relationship between the degree of color divergence and the strength of sexual selection, especially when controlling for neutral genetic distance. We also found a positive correlation between dichromatism and latitude, consistently with the latitudinal pattern of decreasing lineage age and with ecological gradients. Furthermore, a redundancy analyses recovered signals of genetic associations for latitude of breeding range and sexual dichromatism with a subset of SNP outliers, potentially under divergent selection. These results suggest that the joint effects of sexual and ecological selection may have played a role in the radiation of the northern juncos, and support the relevance of sexual selection at early stages of the speciation process.

Key words: sexual dimorphism, sexual selection, plumage coloration, natural selection, phenotypic divergence, speciation, postglacial expansion

Introduction

Sexual selection has long been considered a significant driver of evolutionary diversification and speciation (Darwin 1871; Lande 1981; West-Eberhard 1983; Barraclough et al. 1995; Panhuis et al. 2001). However, the specific role of sexual selection in promoting phenotypic differentiation and lineage divergence remains controversial (Ritchie 2007; Seddon et al. 2008; Kraaijeveld et al. 2011; Seddon et al. 2013). A particular mechanism of speciation by sexual selection has been proposed to operate through the acceleration of the rate of phenotypic change, which may in turn promote differences among allopatric populations in sexually selected traits involved in mate recognition (Price 1998; Seddon et al. 2013; Rowe et al. 2015). This process can lead to fast phenotypic differentiation (Panhuis et al. 2001), and might be especially relevant at early stages of the speciation process (Ritchie 2007; Seddon et al. 2008; Kraaijeveld et al. 2011). Indeed, several cases of highly variable secondary sexual traits in recently radiated systems have been documented, suggesting that sexual selection may be accountable for much of the variation among closely related species in very different taxonomic groups, including spiders (Masta and Maddison 2002), frogs (Boul et al. 2007) and birds (Young et al. 1994; Seddon et al. 2013; Safran et al. 2016; Wilkins et al. 2016).

Rapid divergence among isolated populations driven by sexual selection can be caused initially by random changes (drift) in sexually selected traits and the coevolution of correlated mate preferences, leading to differences in mating success through so-called 'runaway selection' (Fisher 1930; West-Eberhard 1983; Questiau 1999). However, sexual signals necessarily interact with the environmental background and evolve in an ecological context, so that population divergence may be the result of the combined effects of sexual and natural selection (van Doorn et al. 2009; Maan and Seehausen 2011; Butlin et al. 2012; Seehausen et al. 2014). Indeed, the combination of ecological opportunity and sexual selection has been invoked to explain lineage formation in the early stages of speciation and in recent adaptive radiations (e.g. Wagner et al. 2012; Scordato et al. 2014). Correlations of sexual selection with ecological parameters like latitude, habitat type, or migratory behavior have also been reported (Fitzpatrick 1994; Price 1998; Friedman et al. 2009; for review see Badyaev and Hill 2003), lending support to the hypothesis of

sexual and ecological factors jointly driving lineage divergence. However, our understanding of the complex interactions and relative contributions of sexual and natural selection to the diversification process is still limited (Maan and Seehausen 2011; Safran et al. 2016).

Studies of sexually and ecologically selected traits in radiations that include lineages of different ages are particularly useful for gaining insight into the relative roles of sexual and ecological selection in driving lineage diversification (Badyaev and Hill 2003; Kraaijeveld et al. 2011). By studying the evolution of sexual selection across different phylogenetic levels, we can infer how it covaries with the rate of phenotypic divergence in secondary sexual traits, and test for a causal correlation between the strength of sexual selection and the degree of phenotypic differentiation among lineages. Furthermore, we can compare rates of divergence in ecomorphological and sexually selected traits to assess the relative contributions of sexual and natural divergence to the process of lineage differentiation (Arnegard et al. 2010; Safran et al. 2013; Martin and Mendelson 2014). Knowing the evolutionary history of biological systems presenting different spatial settings and occupying distinct environments also allows studying the evolution of sexual selection in relation to the demographic history or the colonization of new habitats (Endler 1980; Price et al. 2008; Wagner et al. 2012). Importantly, phylogenetically informed analyses in systems with lineages of distinct ages allow testing for evidence of positive correlation between sexual selection and the number of cladogenetic events, and infer at what stages of the lineage formation process may be more relevant (Barraclough et al. 1995; Seddon et al. 2008; Maia et al. 2013).

Systems encompassing both old and recently radiated lineages showing variation in ecological and sexually selected traits may be found in the groups that underwent range expansions and colonized new areas across latitudinal gradients following glacial periods (Schluter 2000; Coyne and Orr 2004). Ecological adaptations in some of these systems are accompanied by latitudinal variation in potential sexually selected traits (*e.g.* the *Passerina* buntings; Parulids) suggesting concomitant effects of natural and sexual selection. One such system is provided by the songbird genus *Junco*, which includes highly divergent phylogenetic lineages in Central America as

well as recently diversified lineages ranging from northern Mexico to Canada (Fig. 2.1A; Milá et al. 2007; Chapter I; Friis et al. 2016). The marked differentiation in plumage pattern and color among the recently radiated northern forms of junco suggests that sexual selection may have played a relevant role in their phenotypic diversification. Besides conspicuous variability in sexually selected characters, signals of local adaptation due to climatic factors driving divergence among northern populations of junco have been also detected (see Chapter III). Miller (1941) documented a latitudinal increase in the degree of sexual dichromatism, a relatively common pattern in other avian species that suggests potential interactions between sexual selection and ecological selective pressures related to northern habitats (e.g. Fitzpatrick 1994; Price 1998; Friedman et al. 2009; for review see Badyaev and Hill 2003).

Here, we study patterns of genetic and phenotypic differentiation in the genus *Junco*, including old Central American species and recently radiated North American lineages, and infer the relative roles of sexual selection and ecological factors in driving diversification. We first study the general patterns of neutral genetic structure in the recently radiated northern junco lineages using genomewide single nucleotide polymorphisms (SNPs) obtained by genotyping-by-sequencing (GBS, Elshire et al. 2011). Then we used morphometric and spectrophotometric data from museum specimens in several analytical procedures with four major aims: (i) to compare rates of phenotypic evolution in both traits of ecological importance and plumage coloration by means of discriminant function analyses (DFA) to assess the relative contributions of ecological and sexual selection in the *Junco* radiation; (ii) to quantitatively test the latitudinal pattern of sexual dichromatism proposed by Miller (1941) across the entire distribution of the genus using multivariate and linear regression analyses; (iii) to study the correlation between the degree of divergence in sexually selected characters and a synthetic, comparable index of sexual dichromatism to test the role of sexual selection in driving diversification by means of simple and partial Mantel tests; and (iv) test the potential differential selection effects of ecology and mate choice acting together by testing for genetic associations in a subset of genomewide SNP outliers with both latitude and sexual dichromatism.

Materials and methods

Genotyping-by-sequencing

Here we extended our GBS sample dataset for a total of 243 individual genotypes belonging to the following taxa (with sample sizes in parentheses): *hyemalis* (14), *carolinensis* (22), *aikeni* (12), *mearnsi* (12), *oreganus* (16), *thurberi* (34), *caniceps* (69), *dorsalis* (48), *palliatus* (8) and *phaeonotus* (8) (Fig. 2.1A, Table 2.1).

Genetic structure analyses

To explore genome-wide population structure among recently diverged junco forms, we ran a STRUCTURE (Pritchard et al. 2000) analysis based on SNP data. Using VCFTOOLS, we retained the 8 samples of each population with the lower proportion of missing sites for a final number of 80 samples (Table 2.1). We constructed a data matrix of biallelic SNPs excluding those out of a range of coverage between 2 and 100 or with a genotyping phred quality score below 70. Positions with less than 90% of individuals genotyped were removed from the data matrix, along with those presenting a minor allele frequency (MAF) below 0.02. We implemented a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci. We filtered out SNPs putatively under selection using BayeScan (Foll and Gaggiotti 2008). We used BayeScan with default settings and a thinning interval size of 100 to ensure convergence. For each SNP we obtained the posterior probability for the selection model and the F_{ST} coefficient averaged over populations. For outlier detection and exclusion, we implemented a false discovery rate of 0.1, which resulted in the exclusion of 130 SNPs potentially under selection. To filter out the SNPs under linkage disequilibrium (LD) we used the function snpgdsLDpruning from the SNPrelate package (Zheng 2012) in R Studio (R Studio Team 2015) version 1.0.136 with R (R Core Team 2015) version 3.2.2. We applied the correlation coefficient method with a threshold of 0.2 (method ="corr", ld.threshold=0.2), resulting in a final data matrix of 23,703 SNPs. We converted the *vcf* file to STRUCTURE format using PGDspider (Lischer and Excoffier 2012) version

Species	Genomic analyses		Phenotypic analyses		Latitude
	Complete sequenced dataset (GBS)	STRUCTURE, PCA, F_{ST} and Nei, Mantel tests, redundancy analyses	Per form DFAs	Per sex DFA, linear regression, Mantel tests, redundancy analyses	N (GBIF) No duplicates
vulcani	-	-	-	17 females and 19 males	95 (95) 25
bairdi	-	-	-	15 females and 21males	72 (72) 15
alticola	-	-	-	18 females and 23 males	33 (33) 20
insularis	-	-	-	12 females and 18 males	12 (10) 11
fulvescens	-	-	-	13 females and 22 males	7 (4) 5
phaeonotus	8	8	18 males	17 females and 18 males	53 (53) 50
palliatus	8	8	18 males	18 females and 18 males	22 (20) 4
dorsalis	48	8	15 males	15 females and 15 males	39 (0) 22
caniceps	69	8	24 males	10 females and 24 males	71 (0) 41
thurberi	34	8	25 males	20 females and 25 males	64 (26) 50
oreganus	16	8	12 males	11 females and 12 males	4 (4) 4
mearnsi	12	8	26 males	23 females and 26 males	10 (0) 1
aikeni	12	8	13 males	11 females and 13 males	63 (44) 48
carolinensis	22	8	11 males	12 females and 11 males	35 (35) 35
hyemalis	14	8	22 males	21 females and 22 males	47 (35) 47

Table 2.1. Sample size of the different datasets used in the study including sequencing and genomic analyses, multivariate analyses on phenotypic data, and latitude records.

2.0.5.1. Bash scripts to perform the analyses were created with STRAUTO (Chhatre and Emerson 2016) and we ran the program five times per K, for values of K ranging from 1 to 10 after running a preliminary analysis to infer the lambda value. The burn-in was set to 50K iterations and the analysis ran for an additional100K iterations. Similarity scores among runs and graphics were computed with CLUMPAK (Kopelman et al. 2015).

We used the same SNP data matrix to also examine population structure by means of a principal components analysis (PCA). We used the function snpgdsPCA available in SNPrelate to perform the PCA and obtain the eigenvectors to be plotted (see Annex I for R code). Finally, we computed a matrix of pairwise Nei's distances and F_{ST} values from the same SNP dataset used for the PCA and the STRUCTURE analysis using the R-packages adegenet (Jombart 2008) and hierfstat (Goudet et al. 2015), respectively.

Phenotypic data and analysis

For this Chapter, we used morphometric and colorimetric data from 531 of the sampled museum specimens representing all main junco forms (Table 2.1 and Appendix I). As additional step in this set of analyses, we used the function greedy.wilks from the R-package klaR (Weihs et al. 2005) to implement a stepwise forward variable selection to retain the most discriminant variables (see General Methods for used variables) in separating the predefined forms, and examined overall morphological differentiation among them using male data in a DFA after transforming all variables using natural logarithms. Analyses were conducted in R Studio 1.0.136 with R 3.2.2.

Same 531 samples were used to perform the colorimetric analyses. Once we had computed the avian visual model variables (see General Methods for a detailed explanation of used variables), we used the R function boxplots.stats to detect and exclude eleven potentially wrongly measured samples be implementing a highly conservative coefficient of 10, *i.e.* those data measures 10 times higher or lower than the length of the third and fourth interquartile range. We conducted two different tests based on colorimetric variables using discriminant function analysis. First, to

study the patterns of divergence in plumage coloration among young lineages of junco, we ran a DFA based on male data for the northern forms (*phaeonotus*, *palliatus*, *dorsalis*, *caniceps*, *thurberi*, *oreganus*, *mearnsi*, *aikeni*, *carolinensis* and *hyemalis*). We implemented the same stepwise forward variable selection method as for morphological characters available in the klaR package. Second, we computed a synthetic index of the overall differences between females and males to compare the degree of dichromatism among all junco lineages (including *vulcani*, *bairdi*, *alticola*, *insularis* and *fulvescens*). Sexual dichromatism is considered a relatively reliable, common proxy of the intensity of sexual selection in avian taxa (Owens and Hartley 1998; Dunn et al. 2001; Huang and Rabosky 2014; Cooney et al. 2017). To calculate the index we performed a DFA by sex after applying once again the greedy.wilks stepwise forward selection method to retain the most discriminant variables in terms of sex. Because comparisons among scores of different multivariate analysis and datasets are not statistically valid, we did not separate the analysis for different forms, and ran the DFA for the entire sample space (Montgomerie 2006). We then computed the DFA score means of females and males of each form and their 95% confidence intervals (CI) for graphic comparison. We also conducted a linear regression between the degree of averaged dichromatism and mean geographical coordinates of each form. To compute the latitudinal means, we used the geographic locations of our own field sampling, complemented with GBIF accessions for each junco form (Table 2.1, see Annex I for R code).

Finally, to test the relationship between sexual selection and phenotypic diversification, we used the R-package vegan (Oksanen et al. 2016) to run a simple Mantel test (Mantel 1967) between pairwise distances among the centroids when plotting the first two discriminant functions of the DFA for junco forms and the pairwise sum of the scores of the sexual dichromatism index as an estimate of the intensity of sexual selection experienced by the two lineages under comparison (Seddon et al. 2013). We also ran a partial Mantel test (Smouse et al. 1986) to control for neutral genetic divergence, including the matrix of pairwise Nei's distances to be partialed out. Complementarily, we ran a second simple Mantel test to test for correlation between the two independent matrices (sexual dichromatism and

genetic distance). Significance was computed through 9,999 permutations of rows and columns from the DFA matrix.

Adaptive variation association tests

We tested for associations of adaptive variation in the northern lineages with sexual dimorphism and latitude of breeding range, as proxies of sexual selection and ecological selective pressures, respectively, using redundancy analysis (RDA, Van Den Wollenberg 1977; Legendre and Legendre 1998). Because of their high collinearity (see Results), we ran RDA separately for latitude and sexual dichromatism to obtain an ordination over a single explanatory variable (e.g. Lepš and Šmilauer 2003; Borcard et al. 2011) and then performed a variance partition test to assess the degree of overlapping between each variable's explained variance. The response variable was the frequency of the less frequent allele for each one of 255 biallelic SNPs putatively under selection detected by BayeScan when using a FDR of 0.3 (Meirmans 2015; Rellstab et al. 2015), computed over each of the young northern junco forms. The explanatory variables were averaged latitude and sexual dichromatism per form as previously described. We ran the redundancy analyses using the rda function available in the R-package vegan and obtained their statistical significance by a permutation-based procedure with 9,999 permutations, assuming $\alpha = 0.01$. The variance partition analysis was carried out with the varpart R function, also available in vegan (see Annex I for R code).

Results

Neutral genetic structure among northern Junco *forms*

The STRUCTURE analysis of the young junco lineages for two genetic clusters (K = 2) showed a gradual pattern of divergence from the Mexican *J. p. phaeonotus* to the *J. h. hyemalis* of Canada, approximately separating the yellow-eyed from the dark-eyed forms, with *caniceps* and *dorsalis* forms showing intermediate assignment probabilities, in congruence with their geographic positions. The analysis for K = 3 and K = 4 revealed *carolinensis* and *aikeni* as independent clusters, respectively. In the test for five clusters (K = 5), *thurberi* showed considerable divergence from the

rest of populations, yet also intermediate probabilities of belonging to the main dark-eyed cluster (Fig. 2.1B).

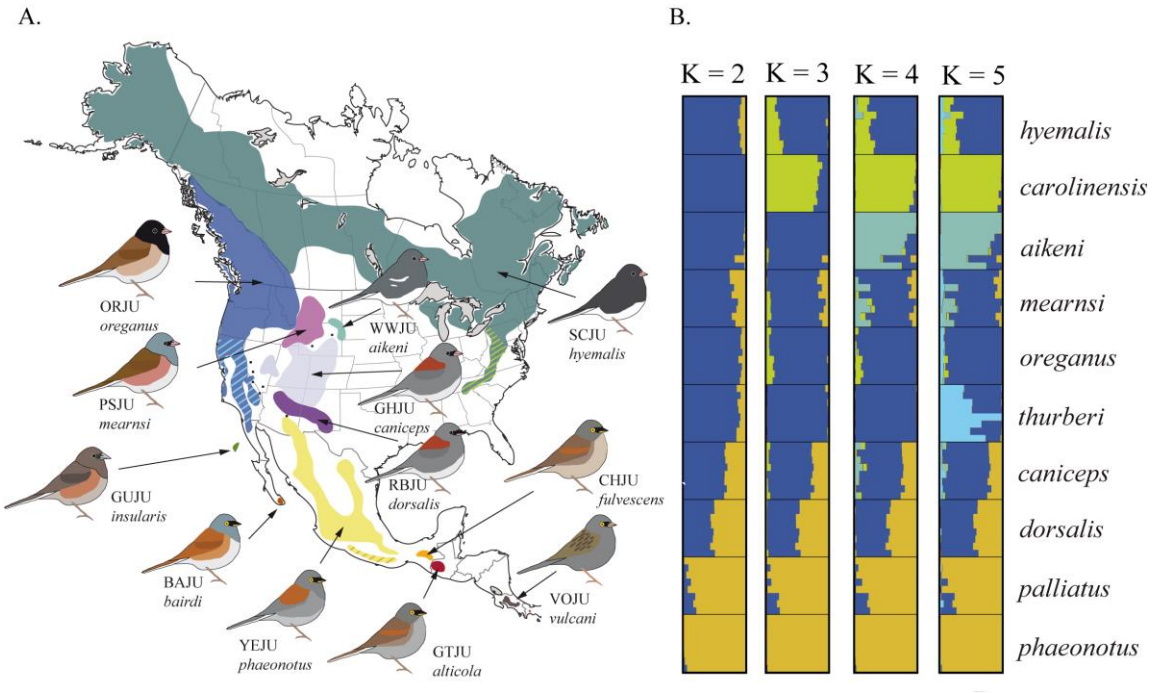


Figure 2.1. Geographic distribution of phenotypic variation the genus *Junco* taxa and neutral genetic structure of the northern forms. (A) Distribution map of the different junco forms. Colored areas correspond to the breeding ranges of the major forms and stripped patterns to subspecific groups (see Table 1 from the General Introduction for a detailed nomenclature). Dots represent isolated localities with hybrid/intermediate individuals and the stripped patterns to subspecific forms *carolinensis* (pistachio green), *phaeonotus* (golden brown) and *thurberi* (light blue). (B) Genetic structure of the northern junco forms from a STRUCTURE analysis based on 23,703 selectively neutral genome-wide SNPs for K = [2-5]. Each horizontal bar corresponds to an individual, with different colors corresponding to posterior assignment probabilities to a given number genetic clusters (K). Colours correspond approximately to those on the range map on Fig. 2.1A.

The PCA yielded similar general patterns. A plot of PC1 (5.5% of explained variance) against PC3 (4% of explained variance against 4.4% of the PC2, but showing better cluster resolution) revealed *carolinensis* and *aikeni* as highly differentiated groups

and clear clustering for all the dark-eyed junco forms. Separation between the *J. phaeonotus* forms was less pronounced, and appeared as close groups to *dorsalis*, the neighbor dark-eyed form from southern USA (Fig. 2.2A).

Nei's distances and F_{ST} values were generally congruent with genetic structure analyses. Southern forms *phaeonotus* and *palliatus* showed the highest values for both indices, while northern forms showed lower levels of pairwise differentiation with a clear increase in the *aikeni* and *carolinensis* forms (Table 2.2).

Nei's\F_{ST}	*phaeonotus*	*palliatus*	*dorsalis*	*caniceps*	*thurberi*	*oreganus*	*mearnsi*	*aikeni*	*carolinensis*	*hyemalis*
phaeonotus		0.075	0.123	0.125	0.094	0.118	0.106	0.113	0.131	0.12
palliatus	0.092		0.088	0.091	0.076	0.089	0.08	0.089	0.102	0.089
dorsalis	0.123	0.072		0.07	0.085	0.08	0.077	0.086	0.09	0.078
caniceps	0.125	0.074	0.043		0.082	0.075	0.073	0.081	0.084	0.072
thurberi	0.108	0.074	0.064	0.061		0.077	0.072	0.079	0.088	0.076
oreganus	0.122	0.076	0.052	0.048	0.059		0.071	0.081	0.083	0.07
mearnsi	0.116	0.073	0.054	0.05	0.06	0.051		0.074	0.083	0.069
aikeni	0.122	0.08	0.059	0.055	0.065	0.058	0.056		0.089	0.078
carolinensis	0.134	0.086	0.057	0.053	0.067	0.054	0.059	0.062		0.078
hyemalis	0.122	0.075	0.05	0.046	0.058	0.046	0.049	0.055	0.05	

Table 2.2. Pairwise Nei's genetic distances (lower diagonal) and F_{ST} values (upper diagonal) for all northern junco forms based on 23,703 independent, selectively neutral SNP loci.

Patterns of phenotypic differentiation and sexual dichromatism

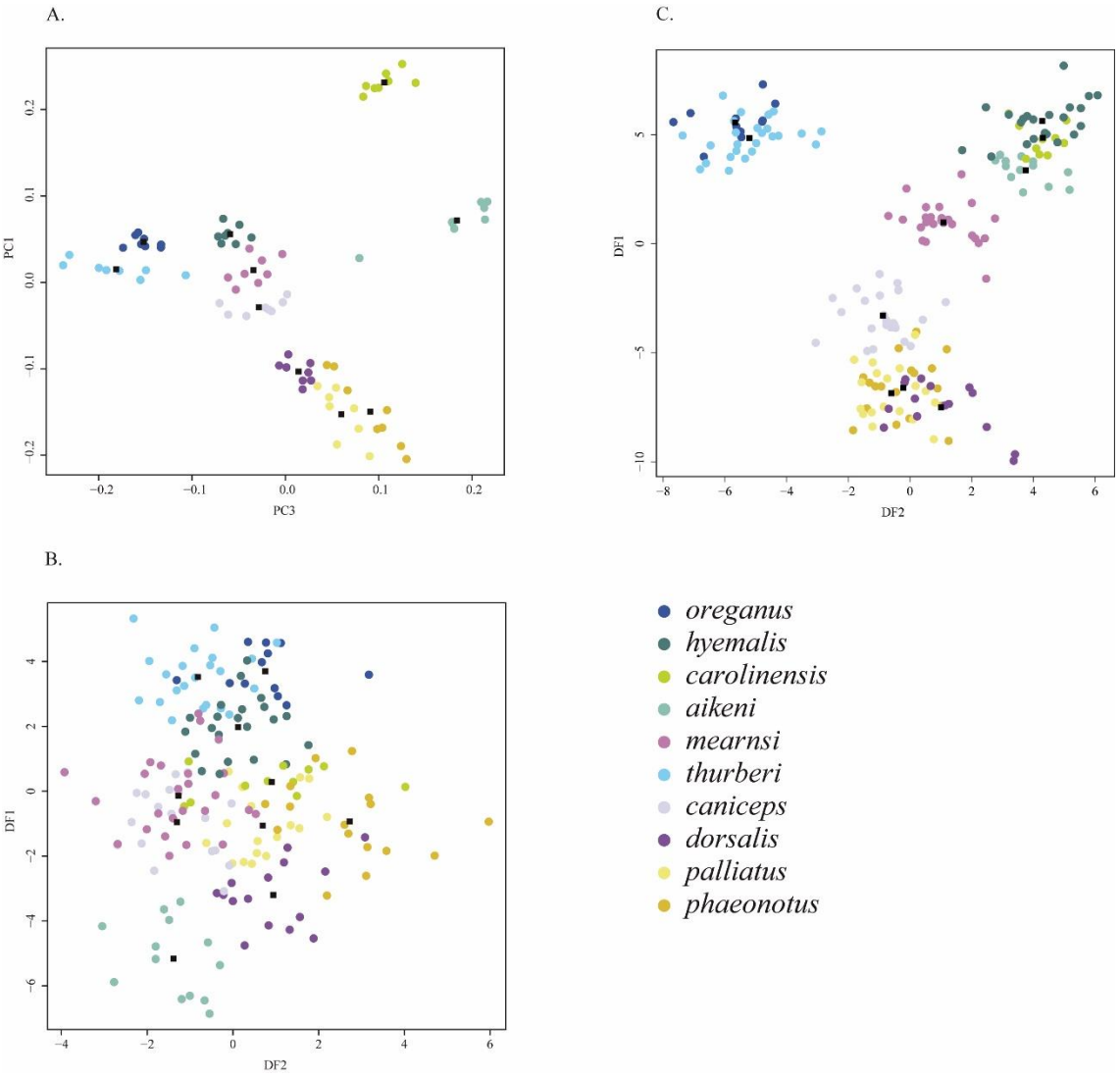
The plot of the first two discriminant functions from a DFA on morphometric variables for the northern lineages of *Junco* revealed a pattern of low clustering among groups (Fig. 2.2B). The forms *aikeni* and *dorsalis*, and to a lesser extent, *thurberi*, *oreganus* and *hyemalis* presented certain degree of separation. The *phaeonotus* centroid showed also a considerable divergence from other northern junco forms, but individuals showed sizable spread across multivariate space. The remaining forms presented extensive overlap.

In contrast, the DFA based on spectral data revealed considerable differentiation in plumage coloration patterns. A plot for the first two discriminant functions showed clear separation of the two black-hooded Oregon junco forms, *oreganus* and *thurberi*, from the rest of lineages, as well as for *mearnsi* and *caniceps*, which occupied more centered positions. The two slate-colored forms, *hyemalis* and *carolinensis* clustered together with the phenotypically similar *aikeni*. Similarly, *phaeonotus* and *palliatus* showed no differentiation between them and overlapped with *dorsalis* (Fig. 2.2C). These patterns were remarkably congruent with the general neutral genetic structure recovered in the PCA (Fig. 2.2A) and with the geographic distribution of the northern juncos (Fig. 2.1A).

The sexual dichromatism index computed from the DFA scores revealed a gradually increasing pattern of differentiation between males and females when ordering the forms from south to north (Fig. 2.3A), with the exception of *caniceps* and *carolinensis*, which did not follow this pattern. The latitudinal signal of increasing dichromatism was also evident when considering only the recently radiated forms, where the yellow-eyed Mexican lineages presented the lowest male-female differentiation values in contrast to the most boreal forms, *hyemalis* and *oreganus*. The linear regression between mean male-female differences and latitude was highly significant ($p = 0.0005$), with latitude explaining 62% of the variance in sexual dichromatism ($R^2 = 0.62$). The Pearson correlation coefficient was equal to 0.79. Remarkably, the pattern persisted within the *oreganus* individuals of our study, with subspecies *thurberi* showing lower dichromatism than northern *oreganus* (Fig. 2.3B).

The simple Mantel test for pairwise color distances and degree of sexual dichromatism revealed a moderate but significant correlation between the two measures (Mantel statistic $r = 0.20$, p-value = 0.04). The correlation and significance increased when controlling for genetic distance in the partial Mantel test, yielding a Mantel statistic r of 0.26 and a p-value of 0.01. In turn, sexual dichromatism and Nei's genetic distance yielded not significant correlation ($r = -0.36$, p-value = 0.85).

Figure 2.2. Neutral genetic structure and phenotypic differences among the recently radiated forms of junco. (A) Genetic structure of northern junco forms based on the first two axis of a principal components analysis of selectively neutral genome-wide SNPs. (B) and (C) show the first two discriminant functions in a discriminant function analysis (DFA) based on morphological variables and plumage color variables, respectively. Marker colors correspond to those on the range map on Fig. 2.1A.



Adaptive variation association tests

Both latitude and sexual dichromatism had significant effects on adaptive genomic variance with p-values equal to 0.001 and 0.019, respectively. Sexual dichromatism

explained 21% of the total adaptive variance (adjusted R^2=0.21), while latitude explained 36% (adjusted R^2=0.36). The RDA scores for latitude as well as dichromatism revealed a pattern of negative correlation with adaptive variance in southern forms of North American juncos (*phaeonotus*, *palliatus*, *dorsalis* and *caniceps*) while more boreal forms showed increasing positive association from south to north, following the phenotypic gradient of sexual dichromatism. Once again, the northernmost form *oreganus* showed the highest correlation scores. In turn, *caniceps* showed low association values, especially in terms of latitude (Table 2.3).

Table 2.3. RDA scores of the constraining latitude and sexual dichromatism variables, explained variance, and p-values. The constrained ordination tests were performed in two separated redundancy analyses, and the statistical significance was computed by a permutation-based procedure with 9,999 permutations, assuming $\alpha = 0.01$.

Species	**Latitude RDA**	**Dichromatism RDA**
phaeonotus	-2.217	-2.430
palliatus	-1.635	-1.864
dorsalis	-0.682	-0.834
caniceps	-0.055	-0.342
thurberi	0.461	0.629
mearnsi	0.480	0.435
aikeni	0.726	0.614
carolinensis	0.858	1.114
hyemalis	0.997	1.178
oreganus	1.066	1.500
	p-value = 0.001	p-value = 0.019
	Adjusted R^2= 0.36	Adjusted R^2 = 0.21

The variance partition analysis revealed a complete lack of orthogonality between the adaptive genetic variance explained by sexual dichromatism and that explained by latitude, i.e. the total of the 21% of the variance explained by sexual dichromatism was also explained by latitude, demonstrating a total overlap between their effects

on adaptive genomic variance. The permutation procedure yielded a p-value equal to 0.003, confirming the significance of the variance fraction explained by both variables. The remaining 15% of variance explained solely by latitude was also significant, with a p-value of 0.011.

Discussion

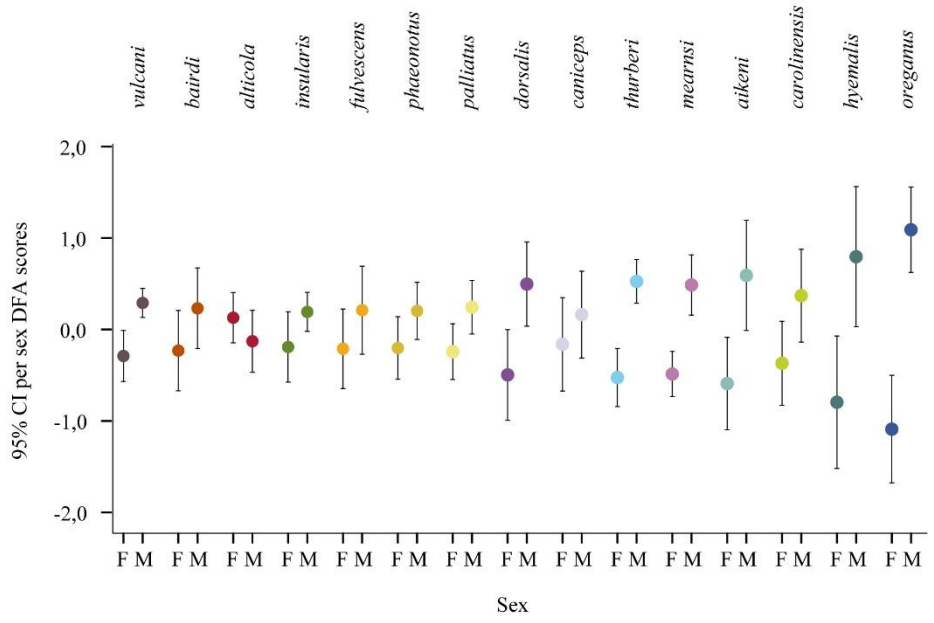
Sexual dichromatism correlates with plumage coloration divergence and latitude

Our results show a strong correspondence between the strength of sexual selection and the degree of phenotypic differentiation in secondary sexual traits across the phylogenetic lineages of the genus *Junco*. Discriminant function analyses on colorimetric variables recovered a clear signal of plumage color differentiation for the northern, recently radiated lineages of junco, as previously reported in a similar analysis of the entire genus (Chapter I; Friis et al. 2016). Interestingly, the DFA of the northern lineages revealed a pattern highly congruent with the neutral genetic structure inferred in the STRUCTURE analysis and especially in the PCA based on neutral genomewide SNP data. In contrast, the DFA of morphometric variables showed low levels of clustering and high overlap among forms, suggesting weaker evolutionary differential pressures on ecomorphological traits than on traits potentially under sexual selection (Panhuis et al. 2001; Arnegard et al. 2010; Safran et al. 2013; Martin and Mendelson 2014).

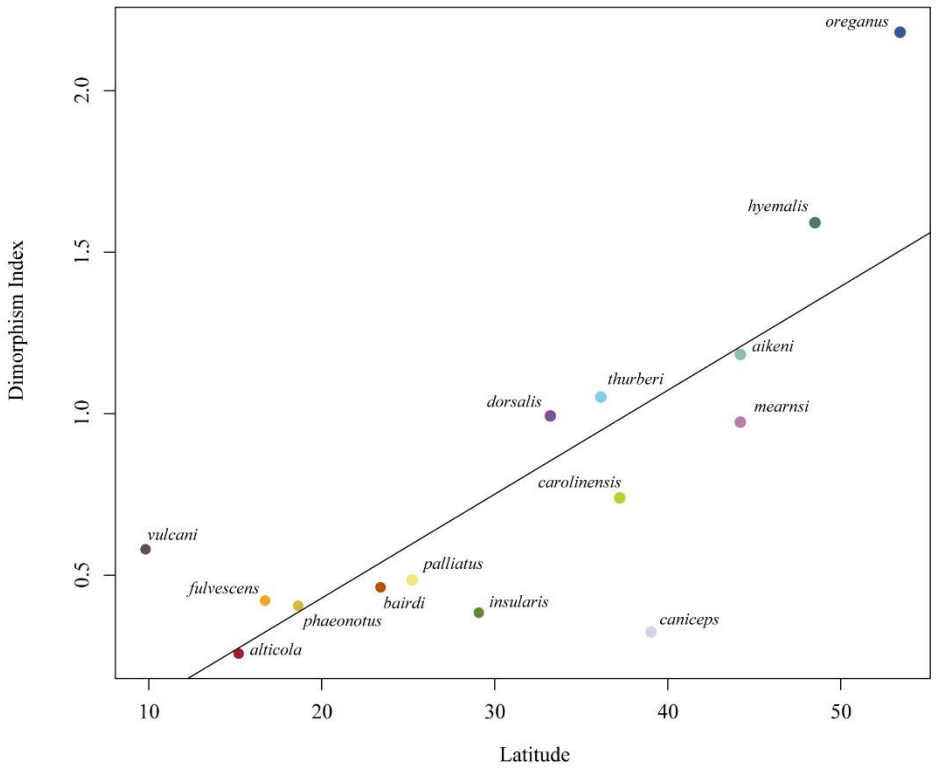
Multivariate and linear regression analyses also confirmed the increasing latitudinal pattern of sexual dichromatism from the divergent Central American lineages to the recently radiated North American forms, as proposed by Miller (1941). Importantly, the latitudinal distribution of the *Junco* species and especially of the postglacial boreal forms reflects not only the ecological gradient across which their demographic expansion occurred, but also the historical sequence of cladogenetic events that resulted in the multiple phylogenetic lineages of the *Junco* complex (Chapter I; Friis et al. 2016). The positive correlation with latitude suggests therefore that sexual dichromatism is a derived, continuous trait that has evolved and increased during the northward recolonization and diversification of the young

Figure 2.3. Latitudinal pattern of gradual increase in sexual dichromatism across *Junco* distribution range. (A) Centered sex-discriminant DFA scores of avian visual model variables (Θ, φ, achieved chroma and normalized brilliance) across the entire sample space for junco forms ordered from south to north. (B) Lineal regression between the degree of averaged dichromatism and mean latitude for each form (p-value = 0.0005, R^2 = 0.62, Pearson correlation coefficient = 0.79).

A.



B.



northern juncos, independently of the changing patterns of plumage coloration themselves.

The significant relationship between plumage color divergence and the degree of sexual dichromatism found in the simple Mantel and especially the partial Mantel test suggests that sexual selection may have had a major role in driving phenotypic divergence among northern junco lineages. Congruently with the higher color similarity among phylogenetically more closely related forms of northern junco (e.g. the yellow-eyed forms *phaeonotus* and *palliatus*; the rufous back forms *dorsalis* and *caniceps*; the black-hooded Oregon forms *thurberi* and *oreganus*; or the slate-colored forms *hyemalis*, *carolinensis* and *aikeni*) the correlation increased when correcting for genetic distance, supporting the existence of divergence driven by sexual selection even when comparing the most recently separated lineages. Mantel and particularly partial Mantel tests have been criticized because the permutation procedure may be an inadequate statistical significance estimator (Raufaste and Rousset 2001). However, partial Mantel tests are deemed suitable when there is low correlation between the independent variables (Castellano and Balletto 2002) as is the case in our study, and under specific assumptions, they are still used in biological analyses (e.g. Funk et al. 2011; Bagley et al. 2017).

Interactions between sexual and natural selection

The redundancy analyses recovered signals of genetic associations for latitude and sexual dichromatism with 255 BayeScan SNP outliers, suggesting the role of both sexual selection and ecological aspects related with latitude in shaping genome-wide adaptive variability in postglacial junco forms. The ordination analyses revealed that up to 36% of the variation in adaptive variability is explained by latitude (p-value = 0.001) and consistently, the ordination scores present an association pattern that increased with latitude, with more extreme forms across the range showing the highest absolute values of correlation. Congruently with the relationship between latitude and the extent of male-female color differentiation, a similar pattern was recovered from the corresponding scores of the ordination analysis over sexual dichromatism. The percentage of explained adaptive genetic variance and the significance were lower in this case, 21% and a p-value of 0.019

respectively. This pattern contrasts with the DFAs on morphological and colorimetric traits that revealed a greater divergence in secondary sexual characters than in ecologically relevant morphometric traits. In addition, the variance partition analysis yielded a complete overlapping between the variance explained by sexual dichromatism and latitude, proving impossible to distinguish between sexual selection and latitude effects for 21% of the variance, while the remaining 15% would correspond solely to latitude effects on explaining adaptive variability. The lack of orthogonality between latitude and dichromatism in explaining the variability on SNPs potentially under selection suggests that adaptive variation explained by sexual selection is also latitudinal-dependent. In other words, the fact that the total variance explained by sexual dichromatism can be explained also by latitude may reflect that adaptive variation due to sexual selection is also structured in terms of variation across the adaptive, latitudinal axis (Lasky et al. 2012), suggesting that sexual and ecological selection may have been coupled processes in the diversification of the northern junco lineages (Butlin et al. 2012).

Association with latitude of breeding is a well-documented, relatively common pattern for sexual dichromatism, yet whether this is due to ecological factors or to non-ecological geographic variation in sexual selection remains controversial (Badyaev and Hill 2003). A similar relationship stands for migratory behavior (e.g. Friedman et al. 2009), arguably because dimorphism facilitates mate recognition and choice during shorter breeding seasons, because rapid establishment of territories increases male-male competition and intrasexual selection, or because ornamentation may be a honest signal of better performance during long seasonal movements (Hamilton 1961; Fitzpatrick 1994; but see Dunn et al. 2015). Other proposed interactions between sexual and natural selection refers to environmental constrains in the production and perception of sexual signals (Maan and Seehausen 2011) also referred as 'external', against 'internal' interactions in which ecologically adaptive traits are also sexually selected, either directly or by linked selection (Safran et al. 2013; Scordato et al. 2014). Dunn et al. (2015) recently proposed that bird coloration may be the result of the simultaneous influence of natural and sexual selection effects on two different axis, the former acting on the type of color and the latter driving male-female differences.

The overlapping signal of association of latitude and dichromatism with adaptive variance in juncos may respond to such hypothesis of migration-related adaptive advantages of sexual dichromatism, considering that migration behavior covaries with latitude. As seasonal movements increased in northernmost populations, natural selection may have favored mate preference behavior across the different lineages of junco because of the potential benefits of mating with a better quality male. The juncos present eumelanin and phaeomelanin-based plumage coloration, and these pigments have been shown to represent honest signals of fitness in some avian systems (Roulin et al. 2008; Safran et al. 2008; Maguire and Safran 2010; Scordato and Safran 2014). The gain and loss of mate preference behaviors based on such signaling to cope with the selective pressures of long distance migration is consistent with the pattern observed in the form *carolinensis*, the non-migratory subspecies of slate-colored junco from the Appalachian Mountains. In contrast to the rest of the slate-colored forms, which usually migrate long distances south of the breeding areas, the seasonal movements of *carolinensis* individuals are mainly altitudinal (Miller 1941; Nolan et al. 2002). The lesser degree of dichromatism observed in this form may reflect a relaxation of sexual selection due to sedentary habits, resulting in a reduction of male-male competition and a return to monochromatism (Fitzpatrick 1994; Badyaev and Hill 2003; Dunn et al. 2015). There is previous evidence of a reduction in sexually selected traits in juncos when shifting from migrant to sedentary habits. In her study from 2004, Yeh reported a decrease of a 22% in the amount of white in tail feathers in a recently established population of *thurberi* that colonized the University of California San Diego (UCSD) campus and became year-round resident. The amount of white in tail feathers has been demonstrated to be involved in mate choice, and to correlate with fitness traits like body size (Hill et al. 1999; McGlothlin et al. 2005), suggesting potential adaptive interactions between natural and sexual selection through honest signaling. Recently, a clear pattern of genetic structure separating UCSD residents from surrounding migratory populations and wintering individuals has been detected (Fudickar et al. 2017), which is consistent with a process of extremely fast genomewide differentiation driven by adaptation to a novel habitat. Another intriguing case is that of *dorsalis*. This form shows no migratory behavior beyond

descending to lower altitudes within the breeding range (Miller 1941), similarly to *carolinensis*, yet the analyses recovered a signal of relatively high degree of sexual dichromatism, contrasting with closely related forms like *phaeonotus* and *caniceps*. While in c*arolinensis* monochromatism is seemingly a derived state lead by the loss of migratory behavior, *dorsalis* represents a case of early gain of sexual dimorphism in a phylogenetic context with no clear underlying factors that require further research.

The signals of genetic association recovered in our analyses are robust and congruent with the ecological aspects and the inferred evolution of sexual selection across the phylogenetic lineages of junco. However, there are a number of caveats and limitations in the methods applied here that need to be discussed. Because we do not have per-individual spectral and SNP data, we used population-based average values for colorimetric variables and allele frequencies, which may reduce the power of the analysis. In addition, using a synthetic variable of sexual dichromatism summarizing sex differences for several distinct color variables across different plumage patches entails a simplification of its potentially complex, polygenic genetic basis, and hinders a straightforward interpretation of the inferred association signal with adaptive variance. However, because pleiotropic effects, variability in regulatory regions and linked variants involved in multiple color traits are common in the genetic determination of bird coloration (e.g. Poelstra et al. 2014; Toews et al. 2016; Uy et al. 2016), even across different plumage patches (Campagna et al. 2017), a few loci may yield an overall signal of correlation with a complex synthetic variable like the sexual dichromatism index computed in this study. Still, the high rate of false positives (type I errors) remains a major concern in genetic association analyses. Here we followed a conservative approach by combining methods of outlier detection relying on allele frequencies (BayeScan), with association tests aiming to reduce the rate of false positives due to distorting factors like geographic structure and demographic history (Meirmans 2015; Rellstab et al. 2015).

The role of sexual selection in the early stages of speciation in the Junco *complex*

There are numerous, compelling cases of rapid diversification of sexually selected traits across closely related species and populations (Price 1998; Kraaijeveld et al. 2011), both in birds (e.g. Uy and Borgia 2000; Wilkins et al. 2016; Campagna et al. 2017) and other taxonomic groups (e.g. Dominey 1984; Masta and Maddison 2002; Boul et al. 2007; Butler et al. 2007). This pattern suggests a role for sexual selection in driving phenotypic diversification at early stages of the speciation process. Several studies have also reported signs of faster evolution in sexually selected traits than in traits of ecological importance (Arnegard et al. 2010; Safran et al. 2013; Martin and Mendelson 2014), reinforcing the argument that sexual selection may contribute to diversification by increasing the rates of phenotypic change in secondary sexual traits across isolated populations (Price 1998; Panhuis et al. 2001).

The recently radiated forms of North American juncos represent one of the most striking examples of rapid phenotypic diversification, having evolved into at least six highly differentiated forms in only 18,000 years c.a. (Milá et al. 2007; Chapter I; Friis et al. 2016). These forms are not differentiated only in their distinctive coloration patterns, but also present considerable genetic structure. This suggests that the current parapatric distribution and the extant hybrid zones among them may be the result of secondary contact zones arisen after the expansion of independent populations that established during the northward postglacial recolonization. Minor divergence has been detected in ecomorphological traits, which fits the hypothesis of divergence arising by an increase of the overall rate of change due to sexual selection acting differentially among genetically divergent junco lineages. Under these assumptions, sexual selection driving fast phenotypic divergence may proceed independently of ecological factors (Panhuis et al. 2001; Kraaijeveld et al. 2011). However, the correlation between sexual dichromatism and latitude and the overlapping association signals of both parameters with the variability of loci putatively under divergent selection found in northern juncos is congruent with the more predominant proposed models of speciation of natural and sexual selection jointly driving diversification (Kraaijeveld et al. 2011; Maan and Seehausen 2011; Butlin et al. 2012; Wagner et al. 2012; Safran et al. 2013). Still, while sexual dichromatism is correlated with latitude, various other distinct

patterns of coloration are not. At the same latitude, we can find highly divergent forms in terms of feather color, a pattern difficult to explain by ecological interactions with mate choice behavior. The process leading to the astounding color diversity of the juncos may be explained in the conceptual framework of mutation-order speciation theory (Schluter 2009; Nosil and Flaxman 2011): during the first stages of the postglacial recolonization process, mutations underlying color changes may have stochastically appeared in isolated populations and been positively selected as indicators of fitness, perhaps of adaptation to long seasonal displacements or local conditions. However, due to lack of gene flow, these traits could become fixed independently in different populations. High evolvability of feather color patterns and the stochastic order in which mutation appears may thereby have fostered the rapid diversification of northern junco lineages even under similar sexual selective pressures (Schluter 2009; Nosil and Flaxman 2011; Mendelson et al. 2014). A similar hypotheses has been proposed by Winger and Bates (2015) for a number of passerine species across the arid Marañon valley of Peru, although over somewhat longer periods of time.

Sexual selection may therefore promote phenotypic diversification, but the extent to which this diversification can result in species formation remains unclear (Ritchie 2007; Kraaijeveld et al. 2011; Seddon et al. 2013). A number of studies have documented a relationship between speciation rate and strength of sexual selection (e.g. Barraclough et al. 1995; Seddon et al. 2008; Kraaijeveld et al. 2011; Maia et al. 2013; Seddon et al. 2013; but see Huang and Rabosky 2014) or changes in the intensity of sexual selection (Gomes et al. 2016). In recently diversified systems, sexual selection may have a predominant role as promoter of premating isolation barriers by accelerating evolutionary divergence in signals involved in species recognition, preventing species to merge upon secondary contact (Price 1998, 2008). Whether this is the case in northern junco lineages is currently difficult to infer. Their genetic distinctiveness and highly divergent patterns of plumage coloration suggest that there may be certain degree of reproductive isolation in some areas, but in others reproductive barriers are clearly absent, and juncos form hybrid zones where parapatric forms come into contact. Estimates of assortative mating and hybrid fitness at these areas of introgression are largely lacking, and will

be necessary to fully understand the degree of reproductive isolation among some junco forms. If, as hypothesized, the present parapatric limits are the result of recent secondary contact after their postglacial diversification in isolated populations, premating barriers to gene flow may have not been sufficiently developed, and the current lineages may fuse if extensive gene flow persists, erasing incipient lineage formation (Grant and Grant 2008; Garrick et al. 2014). Alternatively, contact zones may be stable and ongoing divergence could culminate in a set of fully isolated species, which would yield a positive correlation between sexual selection strength and speciation rate at a phylogenetic level, in agreement with proposed models of speciation by means of the combined effects of sexual selection and local adaptation. In either case, the analyses reported in this study reveal a complex array of sexual and ecological factors as potential drivers of the rapid radiation of the northern lineages of *Junco*, and provide new evidence for the proposed role of sexual selection as a promoter of lineage divergence, especially when interacting with natural selection.

Conclusions

Our analyses confirm the ecological pattern of sexual dichromatism gradually increasing with latitude in the *Junco* system, reinforcing the hypothesis of stronger sexual selection in the North American postglacial lineages. Correlative tests also demonstrated significant dependence between the degree of divergence in terms of plumage coloration and the level of sexual dichromatism, a pattern that contrasted with the lower signal of differentiation in ecomorphological traits, suggesting that sexual selection may have been a predominant evolutionary force in driving phenotypic diversification among recently radiated forms of junco. However, redundancy analyses revealed overlapping effects of both latitude and sexual dichromatism in shaping adaptive variance, suggesting a role for sexual and ecological factors jointly driving lineage differentiation. These results, along with the patterns of neutral genetic structure of the recently radiated lineages of junco, depict a scenario of rapid divergence in isolation at early stages of the speciation process, followed a by a secondary contact phase. Whether or not barriers to reproduction have developed sufficiently to complete the macroevolutionary

lineage formation, the analyses reported here reveal a complex array of sexual and ecological factors as potential drivers of the rapid radiation of the northern juncos, and provide new evidence for the proposed models of lineage divergence promoted by natural and sexual selection.

References

Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. The American Naturalist 176:335-356.

Badyaev, A. V. and G. E. Hill. 2003. Avian sexual dichromatism in relation to phylogeny and ecology. Annual Review of Ecology, Evolution, and Systematics 34:27-49.

Bagley, R. K., V. C. Sousa, M. L. Niemiller, and C. R. Linnen. 2017. History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (Neodiprion lecontei). Molecular ecology 26:1022-1044.

Barraclough, T. G., P. H. Harvey, and S. Nee. 1995. Sexual Selection and Taxonomic Diversity in Passerine Birds. Proceedings of the Royal Society B-Biological Sciences 259:211.

Borcard, D., F. Gillet, and P. Legendre. 2011. Chapter 6. Canonical Ordination. Pp. 153-226. Numerical Ecology with R. Springer.

Boul, K. E., W. C. Funk, C. R. Darst, D. C. Cannatella, and M. J. Ryan. 2007. Sexual selection drives speciation in an Amazonian frog. Proceedings of the Royal Society of London B: Biological Sciences 274:399-406.

Butler, M. A., S. A. Sawyer, and J. B. Losos. 2007. Sexual dimorphism and adaptive radiation in Anolis lizards. Nature 447:202.

Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, C. R. Castillo, W. Diao, M. E. Maan, S. Paolucci, and F. J. Weissing. 2012. What do we need to know about speciation? Trends in Ecology & Evolution 27:27-39.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, and I. J. Lovette. 2017. Repeated divergent selection on pigmentation genes in a rapid finch radiation. Science Advances 3:e1602404.

Castellano, S. and E. Balletto. 2002. Is the partial Mantel test inadequate? Evolution 56:1871-1873.

Chhatre, V. and K. Emerson. 2016. StrAuto: Automation and parallelization of STRUCTURE analysis. See http://strauto. popgen. org.

Cooney, C. R., J. A. Tobias, J. T. Weir, C. A. Botero, and N. Seddon. 2017. Sexual selection, speciation and constraints on geographical range overlap in birds. Ecology Letters.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Darwin, C. 1871. The descent of man, 2 Vols. London 81:130-131.

Dominey, W. J. 1984. Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian Drosophila.

Dunn, P. O., J. K. Armenta, and L. A. Whittingham. 2015. Natural and sexual selection act on different axes of variation in avian plumage color. Science advances 1:e1400155.

Dunn, P. O., L. A. Whittingham, and T. E. Pitcher. 2001. Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. Evolution 55:161-175.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Endler, J. A. 1980. Natural selection on color patterns in Poecilia reticulata. Evolution 34:76-91.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford Univ. Press, Oxford.

Fitzpatrick, S. 1994. Colourful migratory birds: evidence for a mechanism other than parasite resistance for the maintenance of'good genes' sexual selection. Proceedings of the Royal Society of London B: Biological Sciences 257:155-160.

Foll, M. and O. E. Gaggiotti. 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977-993.

Friedman, N. R., C. M. Hofmann, B. Kondo, and K. E. Omland. 2009. Correlated evolution of migration and sexual dichromatism in the New World orioles (Icterus). Evolution 63:3269-3274.

Friis, G., P. Aleixandre, R. Rodríguez-Estrella, A. G. Navarro-Sigüenza, and B. Milá. 2016. Rapid postglacial diversification and long-term stasis within the

songbird genus Junco: phylogeographic and phylogenomic evidence. Molecular Ecology 25:6175-6195.

Funk, D. J., S. P. Egan, and P. Nosil. 2011. Isolation by adaptation in Neochlamisus leaf beetles: host-related selection promotes neutral genomic divergence. Molecular Ecology 20:4671-4682.

Garrick, R. C., E. Benavides, M. A. Russello, C. Hyseni, D. L. Edwards, J. P. Gibbs, W. Tapia, C. Ciofi, and A. Caccone. 2014. Lineage fusion in Galápagos giant tortoises. Molecular ecology 23:5276-5290.

Gomes, A. C. R., M. D. Sorenson, and G. C. Cardoso. 2016. Speciation is associated with changing ornamentation rather than stronger sexual selection. Evolution 70:2823-2838.

Goudet, J., T. Jombart, and M. J. Goudet. 2015. Package 'hierfstat'.

Grant, B. R. and P. R. Grant. 2008. Fission and fusion of Darwin's finches populations. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363:2821-2829.

Hamilton, T. H. 1961. On the functions and causes of sexual dimorphism in breeding plumage characters of North American species of warblers and orioles. The American Naturalist 95:121-123.

Hill, J. A., D. A. Enstrom, E. D. Ketterson, V. J. Nolan, and C. Ziegenfus. 1999. Mate choice based on static versus dynamic secondary sexual traits in the dark-eyed junco. Behavioral Ecology 10:91-96.

Huang, H. and D. Rabosky. 2014. Sexual Selection and Diversification: Reexamining the Correlation between Dichromatism and Speciation Rate in Birds. American Naturalist, The 184:E101-E114.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources 15:1179-1191.

Kraaijeveld, K., F. J. L. Kraaijeveld-Smit, and M. E. Maan. 2011. Sexual selection and speciation: the comparative evidence revisited. Biological Reviews 86:367-377.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the USA 78:3721-3725.

Lasky, J. R., D. L. Des Marais, J. McKAY, J. H. Richards, T. E. Juenger, and T. H. Keitt. 2012. Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Molecular Ecology 21:5512-5529.

Legendre, P. and L. Legendre. 1998. Numerical ecology: second English edition. Developments in environmental modelling 20.

Lepš, J. and P. Šmilauer. 2003. Multivariate analysis of ecological data using CANOCO. Cambridge university press.

Lischer, H. E. L. and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Maan, M. E. and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecology letters 14:591-602.

Maguire, S. E. and R. J. Safran. 2010. Morphological and genetic predictors of parental care in the North American barn swallow Hirundo rustica erythrogaster. Journal of avian biology 41:74-82.

Maia, R., D. R. Rubenstein, and M. D. Shawkey. 2013. Key ornamental innovations facilitate diversification in an avian radiation. Proceedings of the National Academy of Sciences 110:10687-10692.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27:209-220.

Martin, M. D. and T. C. Mendelson. 2014. Changes in sexual signals are greater than changes in ecological traits in a dichromatic group of fishes. Evolution 68:3618-3628.

Masta, S. E. and W. P. Maddison. 2002. Sexual selection driving diversification in jumping spiders. Proceedings of the National Academy of Sciences 99:4442-4447.

McGlothlin, J. W., P. G. Parker, V. Nolan Jr, and E. D. Ketterson. 2005. Correlational selection leads to genetic integration of body size and an attractive plumage trait in dark-eyed juncos. Evolution 59:658-671.

Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to avoid them. Molecular Ecology 24:3223-3231.

Mendelson, T. C., M. D. Martin, and S. M. Flaxman. 2014. Mutation-order divergence by sexual selection: diversification of sexual signals in similar environments as a first step in speciation. Ecology letters 17:1053-1066.

Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society B-Biological Sciences 274:2653-2660.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Montgomerie, R. 2006. Analyzing colors.

Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus, S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Nosil, P. and S. M. Flaxman. 2011. Conditions for mutation-order speciation. Proceedings of the Royal Society of London B: Biological Sciences 278:399-407.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, and R. O'Hara. 2016. Vegan: community ecology package. R Packag. 2.3-5.

Owens, I. P. and I. R. Hartley. 1998. Sexual dimorphism in birds: why are there so many different forms of dimorphism? Proceedings of the Royal Society of London B: Biological Sciences 265:397-407.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends in Ecology and Evolution 16:325-413.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Price, T. 1998. Sexual selection and natural selection in bird speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 353:251-260.

Price, T. 2008. Speciation in birds. Roberts and Company, Greenwood Village, Colorado.

Price, T. D., P. J. Yeh, and B. Harr. 2008. Phenotypic plasticity and the evolution of a socially selected trait following colonization of a novel environment. American Naturalist 172:S49-S62.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Questiau, S. 1999. How can sexual selection promote population divergence? Ethology Ecology & Evolution 11:313-324.

R_Core_Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R_Studio_Team. 2015. RStudio: Integrated Development for R. R Studio, Inc., Boston, MA.

Raufaste, N. and F. Rousset. 2001. Are partial mantel tests adequate? Evolution 55:1703-1705.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24:4348-4370.

Ritchie, M. G. 2007. Sexual selection and speciation. Annual Review of Ecology and Systematics 38:79-102.

Roulin, A., B. Almasi, A. Rossi-Pedruzzi, A.-L. Ducrest, K. Wakamatsu, I. Miksik, J. D. Blount, S. Jenni-Eiermann, and L. Jenni. 2008. Corticosterone mediates the condition-dependent component of melanin-based coloration. Animal Behaviour 75:1351-1358.

Rowe, M., T. Albrecht, E. R. Cramer, A. Johnsen, T. Laskemoen, J. T. Weir, and J. T. Lifjeld. 2015. Postcopulatory sexual selection is associated with accelerated evolution of sperm morphology. Evolution 69:1044-1052.

Safran, R., E. Scordato, M. Wilkins, J. K. Hubbard, B. Jenkins, T. Albrecht, S. Flaxman, H. Karaardıç, Y. Vortman, and A. Lotem. 2016. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. Molecular Ecology 25:3865-3883.

Safran, R. J., J. S. Adelman, K. J. McGraw, and M. Hau. 2008. Sexual signal exaggeration affects physiological state in male barn swallows. Current Biology 18:R461-R462.

Safran, R. J., E. S. Scordato, L. B. Symes, R. L. Rodríguez, and T. C. Mendelson. 2013. Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. Trends in ecology & evolution 28:643-650.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323:737-741.

Scordato, E. S. and R. J. Safran. 2014. Geographic variation in sexual selection and implications for speciation in the Barn Swallow. Avian Research 5:8.

Scordato, E. S., L. B. Symes, T. C. Mendelson, and R. J. Safran. 2014. The role of ecology in speciation by sexual selection: a systematic empirical review. Journal of Heredity 105:782-794.

Seddon, N., C. A. Botero, J. A. Tobias, P. O. Dunn, H. E. MacGregor, D. R. Rubenstein, J. A. C. Uy, J. T. Weir, L. A. Whittingham, and R. J. Safran. 2013. Sexual selection accelerates signal evolution during speciation in birds. Pp. 20131065. Proc. R. Soc. B. The Royal Society.

Seddon, N., R. M. Merrill, and J. A. Tobias. 2008. Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. American Naturalist 171:620-631.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, and e. al. 2014. Genomics and the origin of species. Nature Reviews Genetics 15:176-192.

Smouse, P. E., C. J. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology. 35:627-632.

Toews, D. P., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Current Biology 26:2313-2318.

Uy, J. A. C. and G. Borgia. 2000. Sexual selection drives rapid divergence in bowerbird display traits. Evolution 54:273-278.

Uy, J. A. C., E. A. Cooper, S. Cutie, M. R. Concannon, J. W. Poelstra, R. G. Moyle, and C. E. Filardi. 2016. Mutations in different pigmentation genes are associated

with parallel melanism in island flycatchers. Pp. 20160731. Proc. R. Soc. B. The Royal Society.

Van Den Wollenberg, A. L. 1977. Redundancy analysis an alternative for canonical correlation analysis. Psychometrika 42:207-219.

van Doorn, G. S., P. Edelaar, and F. J. Weissing. 2009. On the origin of species by natural and sexual selection. Science 326:1704-1707.

Wagner, C., L. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366-369.

Weihs, C., U. Ligges, K. Luebke, and N. Raabe. 2005. klaR analyzing German business cycles. Pp. 335-343. Data analysis and decision support. Springer.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. The Quarterly Review of Biology 58:155-183.

Wilkins, M. R., H. Karaardıç, Y. Vortman, T. L. Parchman, T. Albrecht, A. Petrželková, L. Özkan, P. Pap, J. K. Hubbard, and A. K. Hund. 2016. Phenotypic differentiation is associated with divergent sexual selection among closely related barn swallow populations. Journal of Evolutionary Biology 29:2410-2421.

Winger, B. M. and J. M. Bates. 2015. The tempo of trait divergence in geographic isolation: avian speciation across the Marañon Valley of Peru. Evolution 69:772-787.

Yeh, P. J. 2004. Rapid evolution of a sexually selected trait following population establishment in a novel habitat. Evolution 58:166-174.

Young, J. R., J. W. Hupp, J. W. Bradbury, and C. E. Braun. 1994. Phenotypic divergence of secondary sexual traits among sage grouse, Centrocercus urophasianus, populations. Animal behaviour 47:1353-1362.

Zheng, X. 2012. SNPRelate: parrallel computing toolset for genome-wide association studies. R package version 95.

CHAPTER III: Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird

Junco h. oreganus, picture by Roy Hancliff, www.royhancliff.com

Guillermo Friis, Guillermo Fandos, Amanda Zellmer, John McCormack, Brant Faircloth, Borja Milá

Abstract

The formation of independent evolutionary lineages involves neutral and selective factors, and understanding their relative roles in population divergence is a fundamental goal of speciation research. Correlations between allele frequencies and environmental variability across populations can reveal the role of selection in driving divergence, yet the relative contribution of neutral processes like drift in geographic isolation can be difficult to establish. Recently diversified systems with different degrees of geographic isolation across steep ecological gradients provide ideal scenarios to apply genetic-environment association analyses (GEA) while controlling for the effects of population history and structure. The Oregon junco (Aves: Emberizidae) of western North America ranges from northern Baja California through southern Alaska, and it has diversified into several phenotypically distinct forms as a result of a rapid postglacial expansion within just the last 20,000 years. Analysis of over 30,000 SNP loci in 136 individuals from seven Oregon junco taxa revealed marked genetic structure, with differentiated populations in isolated, dry southern mountain ranges, and more admixed, recently expanded populations in humid northern latitudes. We used redundancy analysis (RDA) to study correlations between genomic and environmental variance, and tested for three specific modes of evolutionary divergence among Oregon junco populations, including (i) drift in allopatry or in isolation by distance, (ii) differentiation along continuous selective gradients, and (iii) isolation by adaptation. We found evidence of strong isolation by drift, especially in southern mountains, but also signals of local adaptation in several populations, which were particularly evident when controlling for population history. We identified several putative ecological drivers of local adaptation, including temperature, precipitation, vegetation cover and greenness. A Bayesian outlier analysis identified variants under selection scattered across the genome, with no clear "islands" of high divergence. These results suggest that local adaptation can promote rapid differentiation over short periods when acting over multiple loci across the genome, and support a diversification process driven by multiple selective factors in Oregon juncos.

Key words: speciation, redundancy analysis, isolation by distance, isolation by adaptation, selective gradients, postglacial expansion

Introduction

Lineage diversification involves both selective and neutral factors, and elucidating their relative strengths and interactions in the process of evolutionary divergence is essential to understand the mechanisms underlying the early stages of speciation (Coyne and Orr 2004; Nosil 2012). Divergent natural selection is a fundamental mechanism of lineage differentiation (Darwin 1859; Coyne and Orr 2004), and forms the foundation of the 'ecological speciation' model, a process by which reproductive isolation arises as a by-product of cumulative, ecologically adaptive changes (Mayr 1947; Schluter 2000; Rundle and Nosil 2005). In turn, accumulation of genetic differences caused by drift in geographic isolation or in isolation-by-distance (IBD, Wright 1943; Wright 1946) has been proposed as a mode of divergence driven by neutral factors (Mayr 1954, 1963), a mechanism particularly strong in populations of small effective size (e.g. Carson 1975; Templeton 1981; Uyeda et al. 2009).

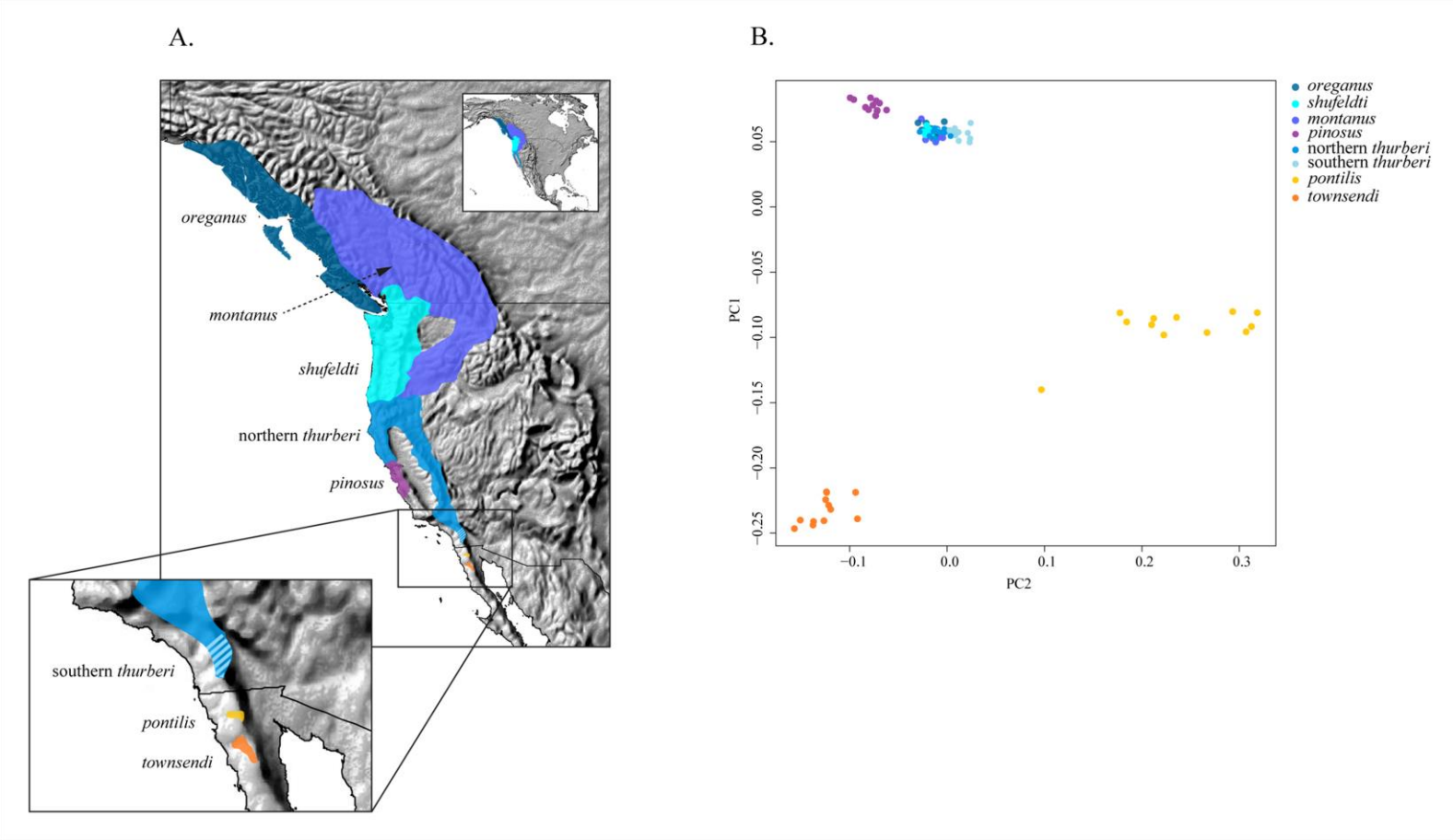
Selection and drift can also act jointly and even interact in a number of ways during evolutionary divergence. Geographic distance usually implies environmental differences that may drive adaptation to local conditions and ecological differentiation, even if populations are connected by moderate gene flow (Schluter 2000; Rundle and Nosil 2005). The evolution of local adaptation can in turn result in isolation-by-adaptation (IBA), a mode of divergence where adaptive changes lead to intrinsic barriers to gene flow, enabling genome-wide differentiation at both neutral and selected loci (Nosil et al. 2008; Funk et al. 2011). Consequently, geographic distance and ecological divergence may promote similar patterns of genetic diversity among populations, and ultimately they both contribute to spatial patterns of biological diversity, so that teasing apart the roles of neutral evolution and ecological adaptation in evolutionary diversification requires approaches that account for both environmental heterogeneity and neutral population structure (Wang and Bradburd 2014; Frichot et al. 2015; Rellstab et al. 2015; Forester et al. 2016).

Our capacity to assess the relative roles of adaptation and neutral differentiation in driving population divergence has benefitted from our increasing ability to survey

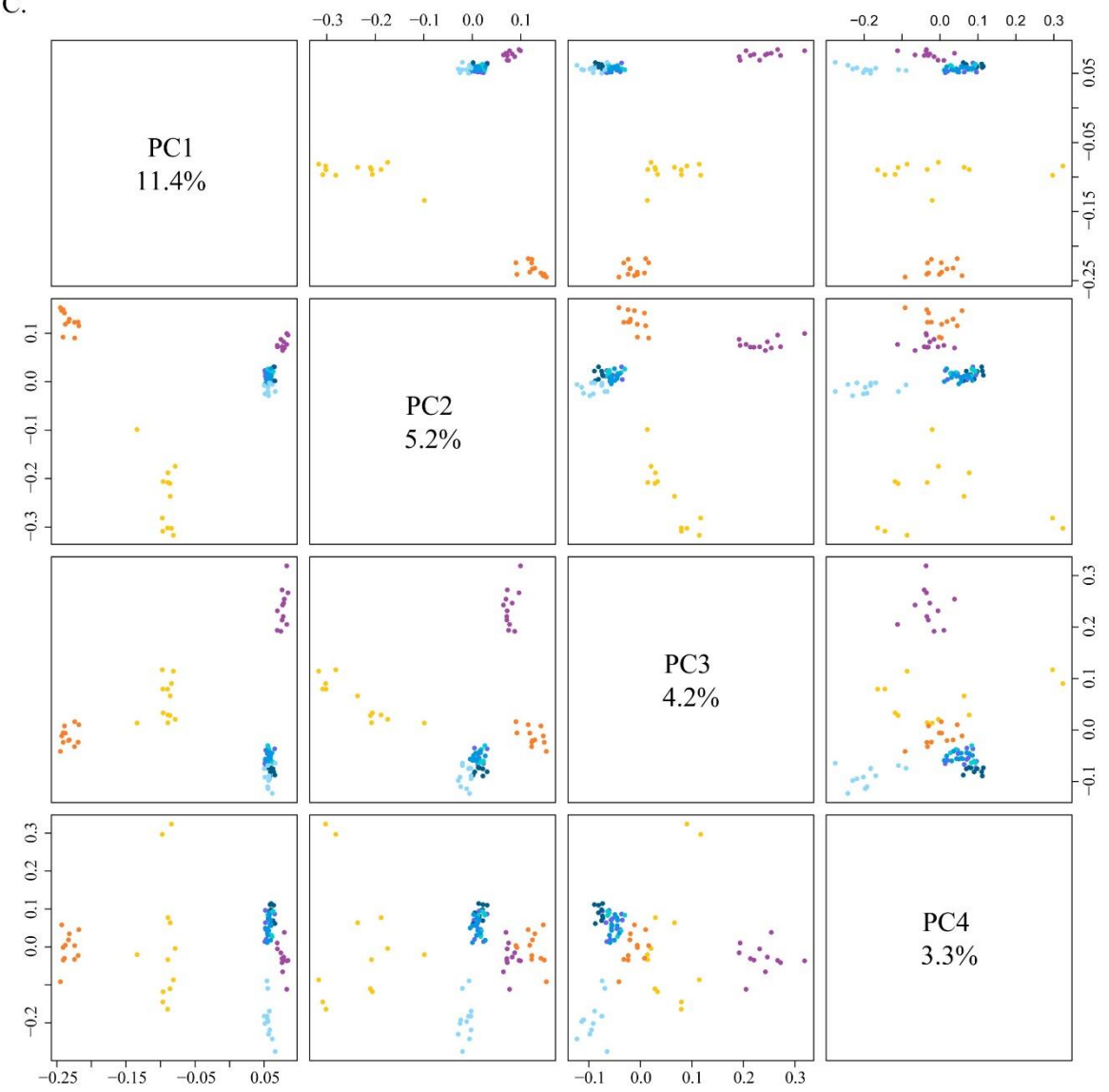
genome-wide variation thanks to the development of next generation sequencing (NGS) techniques (McCormack et al. 2013; Faria et al. 2014). The increasingly large number of loci afforded by NGS provides improved resolution to detect neutral population structure and patterns of gene flow among differentiated lineages. In addition, highly differentiated loci identified as outliers in an F_{ST} distribution can be interpreted as potential targets of divergent selection standing out in a background of balanced or neutrally maintained genomic variation (Faria et al. 2014; Rellstab et al. 2015). Methods of outlier detection relying solely on allele frequencies are sensitive to the confounding effects of historical factors, however, such as past sudden changes in population size or strong drift in small populations that may result in high rates of false positives (Edmonds et al. 2004; Kawecki and Ebert 2004; Billiard et al. 2005; Christmas et al. 2016). Moreover, changes in allele frequencies due to local adaptation are sometimes undetected by outlier analyses (Pritchard and Di Rienzo 2010; Bierne et al. 2011; Rellstab et al. 2015). Alternative approaches that integrate environmental parameters by identifying allele frequencies that correlate with ecological variability have proven useful to detect signals of adaptation, especially when selective forces are weak (Frichot et al. 2015; Rellstab et al. 2015). These methods, known as genetic-environment association (GEA, Hedrick et al. 1976; Mitton et al. 1977) analyses, have the potential to reveal genetic patterns of differentiation due to local adaptation while testing for the role of specific environmental variables as drivers of selection. Importantly, GEA methods can correct for population history by controlling for general patterns of neutral genomic variation (Rellstab et al. 2015; Forester et al. 2016), allowing us to separate the respective effects of drift and selection in generating and maintaining variability. GEA analyses have greatly benefitted from the development of high-throughput sequencing techniques, resulting in a number of studies focusing on the genomic variability associated with environmental parameters in groups as diverse as plants (Lasky et al. 2012; Jones et al. 2013; De Kort et al. 2014; Nadeau et al. 2016; Sork et al. 2016), fungus (Ojeda Alayon et al. 2017) and birds (Manthey and Moyle 2015; Safran et al. 2016; Szulkin et al. 2016; Termignoni-García et al. 2017).

Recently diversified systems provide an ideal scenario for studying the relative roles of selective and neutral factors in incipient divergence and speciation. Specifically,

Figure 3.1. Geographic distribution and neutral genetic structure of the Oregon junco forms. (A) Breeding ranges of all the Oregon junco forms. (B) and (C) Genetic structure of Oregon junco forms based on a principal components analysis of independent, selectively neutral genome-wide SNPs. (B) Plot of the PC1 against the PC2. (C) Pairwise plots of the first four principal components.



C.



GEA methods are particularly suitable when the system under study (i) is composed of closely related populations, among which the signals of selection are still recent and detectable; (ii) includes broad geographic distributions encompassing heterogeneous habitats across ecological clines (*i.e.* selective gradients) but also spatially discontinuous habitats so that adaptive and neutral divergence can be assessed in different spatial scenarios; (iii) shows large variability in the degree of geographical isolation among populations, from extensive gene flow to total isolation; and (iv) presents low variability in secondary sexual traits so that differential sexual selection can be ruled out as a major driver of population divergence.

The Oregon junco complex (*Junco hyemalis oreganus*) of western North America provides a particularly well suited system to carry out genome-environment association analysis. The complex originated recently as part of the postglacial radiation of dark-eyed juncos across North America following a northward recolonization of the continent as ice sheets retreated after the last glacial maximum, c.a. 18,000 years ago (Milá et al. 2007; Friis et al. 2016; Milá et al. 2016). Among dark-eyed junco forms, the Oregon junco group presents the highest variability in terms of phenotype and ecological range, encompassing a broad latitudinal range from Baja California to Alaska. All forms of the Oregon junco share a characteristic dark hood, yet there is considerable population variation in plumage color, mainly of the hood, dorsum and flanks, and the complex has been traditionally divided into at least 7 subspecific forms (Dwight 1918; Miller 1941; Nolan et al. 2002), which include, from south to north: *townsendi*, from the San Pedro Mártir mountains in northern Baja California, Mexico; *pontilis*, distributed just north of *townsendi* in the Sierra Juárez mountains, also in Baja California, Mexico; *thurberi*, from the mountains of southern California and Sierra Nevada; *pinosus*, a coastal form from central California, predominantly distributed in the Santa Cruz mountains; *montanus*, distributed across the interior of Oregon, Washington and British Columbia; *shufeldti*, a more coastal form from Oregon and Washington; and *oreganus* from coastal British Columbia and southern Alaska (Miller 1941; Nolan et al. 2002; Fig. 3.1A, Table 3.1).

The diverse spatial configuration of populations and environmental variability across the Oregon junco distribution are critical aspects that will affect our capacity to disentangle the roles of adaptive and neutral factors in explaining genomic variance. Here we use a conceptual framework to classify into three main settings the distinct spatial scenarios observable in the Oregon junco with respect to gene flow and environmental variation. These include (i) geographically isolated populations in similar habitats, as in the case of the Baja California *townsendi* and *pontilis* forms, where low levels of local adaptation and low rates of gene flow should result in limited adaptive divergence and high neutral divergence by drift (Fig. 3.2A); (ii) parapatric populations under divergent ecological conditions, as exemplified by the *pinosus* and *thurberi* forms in California, where divergence is

expected to increase due to local adaptation, while geographic proximity and moderate gene flow should lead to intermediate levels of neutral differentiation by drift (Fig. 3.2B); and (iii) populations found along a continuous environmental gradient, as in the northernmost forms of Oregon junco (*thurberi*, *shufeldti*, *montanus* and *oreganus*) where neutral divergence is expected to be low due to high levels of gene flow, while local adaptation along the gradient may result in a pattern of high differentiation in adaptive variation (Fig. 3.2C).

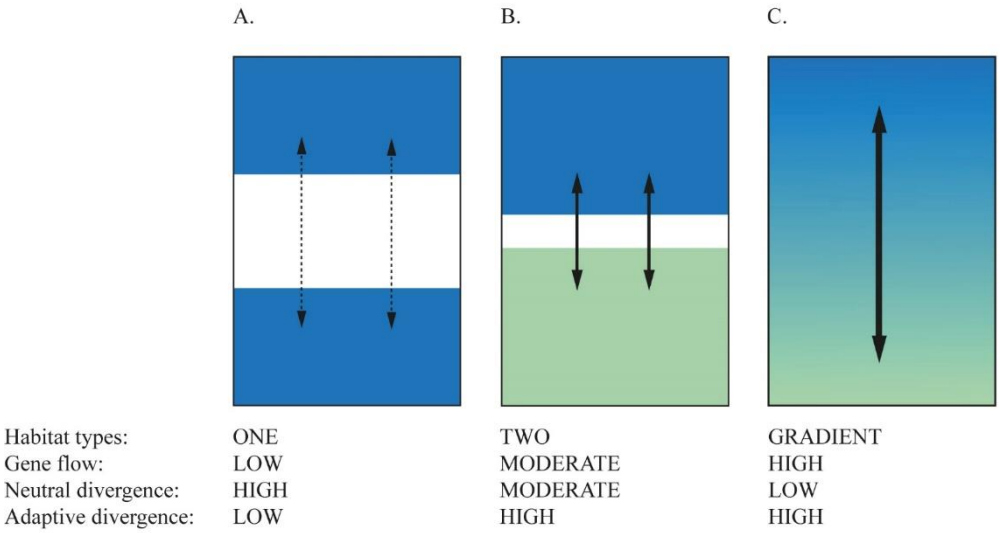


Figure 3.2. Expectations for neutral and adaptive divergence under different environmental and spatial configurations found across the Oregon junco distribution. (A) Geographically isolated populations in similar habitats. (B) Parapatric populations in ecologically divergent habitats. (C) Population continuum across a selective gradient.

Here we use the Oregon junco complex to study how geographic isolation, population history, and local ecological adaptation have driven population differentiation across the range, using extensive population sampling and genome-wide SNPs obtained from 'genotyping by sequencing' (GBS, Elshire et al. 2011). A draft consensus genome of junco has also been sequenced and assembled to be used as a reference. We first assess patterns of neutral genetic structure across the complex using selectively neutral SNPs, and we then look for correlations between

environmental variables and allele frequencies across the Oregon junco distribution using redundancy analysis (RDA). In order to assess niche divergence among selected pairs of taxa, we also use climatic variables to generate niche models and test for significant niche divergence while controlling for spatial autocorrelation. Finally, we map GBS sequences harboring significant outlier loci to the zebra finch (*Taeniopygia guttata*) reference genome in order to recover the chromosomal position of polymorphic sites and explore how adaptive variation is distributed across the genome.

Table 3.1. Oregon junco forms and number of genotyped individuals per locality. State abbreviations are the following: British Columbia (BC) in Canada; Oregon (OR), California (CA) in the USA; Baja California Norte (BCN) in Mexico.

Form	*State*	*Localities*	*Sequenced*
oreganus	BC	Banks Island, Porcher Island, Susan Island	16
shufeldti	OR	Willamette N.F.	12
montanus	OR	Wallowa N.F.	21
n. *thurberi*	CA	Tahoe	18
s. *thurberi*	CA	Mount Laguna	20
pinosus	CA	Santa Cruz Mountains	17
pontilis	BCN	Sierra Juárez	16
townsendi	BCN	Sierra San Pedro Mártir	16
Total			136

Materials and methods

Genotyping-by-sequencing

We used genotyping-by-sequencing (Elshire *et al.* 2011) to obtain individual genotypes from 136 Oregon juncos belonging to the following subspecific taxa: *townsendi* (n=16), *pontilis* (n=16), *thurberi* (n=38), *pinosus* (n=17), *montanus* (n=21), *shufeldti* (n=12), *oreganus* (n=16) (Table 3.1).

Genetic structure analyses

To explore genome-wide population structure among all Oregon junco forms, we ran a principal components analysis (PCA) based on SNP data. Using VCFTOOLS we retained all samples with less than 25% missing data after a 'soft filtering' (coverage range between 2 and 100, minimum phred quality score of 40), resulting in a dataset of 88 samples, and between 8 and 12 individuals per population (24 in the case of *thurberi*). We filtered out all the sites with any non-genotyped individuals or a minor allele frequency (MAF) below 0.02. We also applied a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci, resulting in a matrix of 11,261 variants. We then excluded SNPs putatively under selection using BayeScan (Foll and Gaggiotti 2008) with default settings and a thinning interval size set to 100 to ensure convergence. For each SNP we obtained the posterior probability for the selection model and the F_{ST} coefficient averaged over populations. For outlier detection and exclusion, we implemented a false discovery rate of 0.3. To filter out the SNPs under linkage disequilibrium (LD) we used the function snpgdsLDpruning from the SNPrelate package (Zheng 2012) in R Studio (R Studio Team 2015) version 1.0.136 with R (R Core Team 2015) version 3.2.2. We applied the correlation coefficient method with a threshold of 0.2 (method ="corr", ld.threshold=0.2), resulting in a final data matrix of 9,436 SNPs (Table 3.2). We then used the function snpgdsPCA also available in SNPrelate to perform the PCA and obtain the eigenvectors to be plotted.

We examined population structure by means of the program STRUCTURE (Pritchard et al. 2000), using a smaller, more heavily filtered SNP data matrix to reduce the computational time of the analysis. Using VCFTOOLS, we retained the eight samples of each population (16 in the case of *thurberi*) with the lowest proportion of missing sites for a final number of 64 samples. We constructed a data matrix of biallelic SNPs excluding those out of a range of coverage between 2 and 100, or with a genotyping phred quality score below 70. Positions with less than 90% of individuals genotyped were removed from the data matrix, along with those presenting a MAF below 0.02. Once again, we implemented a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4}, and performed the filtering for non-neutral positions and linkage disequilibrium exactly

as done for the PCA, for a final data matrix of 34,367 SNPs (Table 3.2). We converted the *vcf* file to STRUCTURE format using PGDspider (Lischer and Excoffier 2012) version 2.0.5.1. Bash scripts to perform the analyses were created with STRAUTO (Chhatre and Emerson 2016) and we ran the program five times per K, with K ranging from 1 to 10 after running a preliminary analysis to infer the lambda value. The burn-in was set to 50K iterations and the analysis ran for an additional 100K iterations. Similarity scores among runs and graphics were computed with CLUMPAK (Kopelman et al. 2015).

Finally, to complement the genetic structure analyses, we computed a matrix of pairwise Nei's distances and F_{ST} values from the same SNP dataset used for the PCA using the R-packages adegenet (Jombart 2008) and hierfstat (Goudet et al. 2015), respectively.

Table 3.2. SNP data matrices used in each analyses. General filters included a depth range from 2 to 100 and a p-value for Hardy-Weinberg deviation of 10^{-4}.

Analysis	**Samples**	**Phred score**	**MAF**	**SNPs**	**Missing data**
STRUCTURE	64	70	0.02	34,367	10%
PCA, pairwise F_{ST} and Nei's distance	88	40	0.02	9,436	0%
RDA:					
On all *loci*	88	40	0.02	11,261	0%
On BayeScan outliers	88	40	0.02	87	0%
Genome Scans:					
All lineages	88	40	0.02	29,868	75%
townsendi vs. *pontilis*	24	40	0.05	22,773	75%
townsendi vs. *thurberi*	24	40	0.05	22,516	75%

Redundancy analysis and variance partition

When applying GEA methods there are two main potentially confounding effects related to neutral factors: (i) structure among populations derived from strong drift in isolation may result in genetic patterns similar to those related to adaptive divergence; and (ii) demographic expansions along latitudinal axes may create gradients of allele frequencies at neutral loci correlated with latitude, that in turn would correlate with any environmental variable that changes with latitude, mimicking a pattern of selective sweep and local adaptation (Excoffier and Ray 2008; Excoffier et al. 2009; Rellstab et al. 2015; Forester et al. 2016). Redundancy analysis (Van Den Wollenberg 1977; Legendre and Legendre 1998) is a canonical ordination method that allows computing the variance of a set of response variables that is explained by a number of constraining or explanatory variables. In addition, partial RDA enables computation of this shared variance between two sets of variables while conditioning or holding constant the effects of a third set of covariables. Here we used RDA and partial RDA as implemented in the R package vegan (Oksanen et al. 2016) to explore the associations between genetic variability and environmental data. Ecological data were obtained from 7 of the 19 variables available in the BioClim database (Hijmans et al. 2005), specifically chosen in accordance to their relevance to junco ecology (Miller 1941; Nolan et al. 2002). They measured mean temperature and precipitation over the year (BIO1 and BIO12); mean temperature and precipitation over the warmest quarter (BIO10 and BIO18), which corresponds to the birds' breeding season; isothermality, referring to how the range of day-to-night temperature differs from the range of summer-to-winter, where a value of 100 indicates equality between them; and seasonality of temperature and precipitation (BIO4 and BIO15). We also included three vegetation variables from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellites as available in https://modis.gsfc.nasa.gov: percent tree cover (TREE), Normalized Difference Vegetation Index (NDVI, a measure of canopy greenness), and NDVI's annual standard deviation (std_NDVI). Finally, we included the high-quality elevation data provided by the NASA Shuttle Radar Topographic Mission (SRTM), downloadable from http://www2.jpl.nasa.gov/srtm/ (Table 3.3). All ecological variables were centered and standardized. Following Blanchet et al. (2008), we implemented a forward selection method using the forward.sel function

from the R-package packfor (Dray et al. 2009) to reduce the number of variables in the model. This procedure applies two stopping criteria: a significance level for each tested variable, which we set at 0.01; and a maximum limit for global adjusted R^2, equal to the adjusted R^2 of the RDA model including all initial variables. In doing so we prevent inflation of the overall type I error and of the amount of explained variance. After this, we excluded those retained variables with a variance inflation factor (VIF) over 10 (Borcard et al. 2011) to avoid high colinearity. The specific

Table 3.3. Environmental variables included in the initial stepforward selection method. Significant variables retained by the step forward selection method are shown in bold.

Variable	Description
BIO1	Annual Mean Temperature
BIO3	**Isothermality**
BIO4	Temperature Seasonality (standard deviation *100)
BIO10	**Mean Temperature of Warmest Quarter**
BIO12	Annual Precipitation
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO18	**Precipitation of Warmest Quarter**
NDVI	**Normalized Difference Vegetation Index (greenness)**
Std. NDVI	Greenness seasonality (annual NDVI standard deviation)
TREE	**Tree Cover (%)**
SRTM	**Elevation, NASA Shuttle Radar Topographic Mission**

ecological hypothesis combined with the treatment to select significant variables, reduces the risk of making false inferences due to covariance among ecological parameters not included in the set of constraining variables, unknown to the observer, which may play an important role in adaptive diversification. Despite signs of correlations observable among variables in the partial RDA (especially among BIO18, TREE and NDVI, See Results) we chose not to exclude more variables or to apply dimension-reduction treatments like PCA to the environmental space of variables so as to assess their specific and relative contributions to differentiation

patterns (McCormack et al. 2010) and discuss signals of adaptation with higher confidence. The final selected ecological variables were used as explanatory variables in two RDAs, with and without partialing out the effects of neutral processes, which were approximated by the first two principal components of the PCA of population structure based on selectively neutral loci (see above). As response variables, we used the same SNP dataset used for the PCA, but excluding LD and neutrality filters. SNP data were coded as counts of the alternative allele for each position (*i.e.*, 0, 1 or 2 copies) and transformed following Patterson et al. (2006). Statistical significance was obtained using a permutation-based procedure with 10,000 permutations, assuming $\alpha = 0.01$. We also used variance partitioning to estimate (i) the total proportion of genomic variation explained by ecological variables alone; (ii) by neutral structure alone; and (iii) the effects of both sets of variables. Finally, we repeated the whole RDA treatment for a subset of 87 SNPs identified as selectively divergent by BayeScan, with no conditional treatment. The analyses were conducted in R version 3.2.2 (R code included in Annex I).

Niche divergence tests

To further explore patterns of ecological divergence in the Oregon Junco, we tested for niche divergence applying the method developed by McCormack et al. (2010), a method that allow us to examine each environmental variable separately. To avoid a loss of statistical power due to multiple analyses, we conducted three specific comparisons of forms presenting different patterns of genetic divergence and geographical settings: (i) we compared the *townsendi* and the southern *thurberi* forms, in order to estimate niche divergence between geographically isolated, genetically differentiated forms; (ii) we compared the ecologically divergent *pinosus* with the parapatric northern *thurberi* form, to further test a possible case of isolation-by-adaptation; and (iii) we compared the northern and southern populations of *thurberi*, as conspecific extremes of a potential adaptive gradient (see Results and Discussion sections). We used occurrence points from our own georreferenced field sampling records. The set of occurrence records was further revised to avoid spatial autocorrelation and to match the spatial resolution of environmental variables (1-km grid). Our final dataset comprised 80 localities: *pinosus* (n = 14), *thurberi* north (n = 26), *thurberi* south (n = 19), and *townsendi* (n =

21). We decided to improve quality (geographic accuracy) vs. quantity (number of occurrence records), by using fewer data but with higher spatial accuracy (Engler et al. 2004). To generate a background dataset for each population, we drew 1000 random points from a background representing the geographic range of each junco population. In order to select an appropriately sized area for the niche divergence tests, we included accessible habitats according to the dispersal ability of each population (Soberon and Peterson 2005). We generated background samples from a 100-km "buffer zone" around known occurrences (Warren et al. 2008). For populations with small ranges or small dispersal ability (*thurberi* south, *pontilis* and *townsendi*) we restricted the buffer zone to 10 km to reduce spatial inaccuracies in the null distribution (Barve et al. 2011), after testing different buffer sizes to test the robustness in delimiting accessible areas for juncos. Next, we extracted the environmental data (same as the data used for the RDA, see above) for both occurrence points and random background points from within the geographic range of each junco form. Niche divergence and conservatism was tested by comparing the observed environmental differences among forms against a null model of background divergence (generated by calculating the difference between background points using a bootstrapping approach and 1000 resamples) for each environmental variable using a two-tailed test. We conducted all the analysis using R 3.2.2.

Genome scans

We performed genome scans for different Oregon junco forms using BayeScan.

In order to obtain the chromosomic positions of the SNPs, we mapped the GBS reads against the zebra finch (*Taeniopygia guttata*) genome v87 available in Ensembl (Yates et al. 2016), applying the same set of tools and parameters as for mapping against the junco genome. Using the same set of samples as in the PCA and the RDA, we conducted the analysis for all forms together; for *townsendi* against *pontilis*; and for *townsendi* against all *thurberi* (see Table 3.2 for final dataset sizes). For each of these matrices, we retained only biallelic SNPs with coverage between 2 and 100 and a genotyping phred quality score over 40. Positions with less than 25% of the individuals genotyped were removed from each data matrix, along with those presenting a MAF below 0.05. Once again, we implemented a p-value threshold for

HWE of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci. We ran BayeScan with the same settings used for filtering out SNPs under selection in population structure analyses, but implemented a more conservative 10% FDR for outlier detection. Genome scan plots were conducted in R 3.2.2 using the package qqman (Turner 2014).

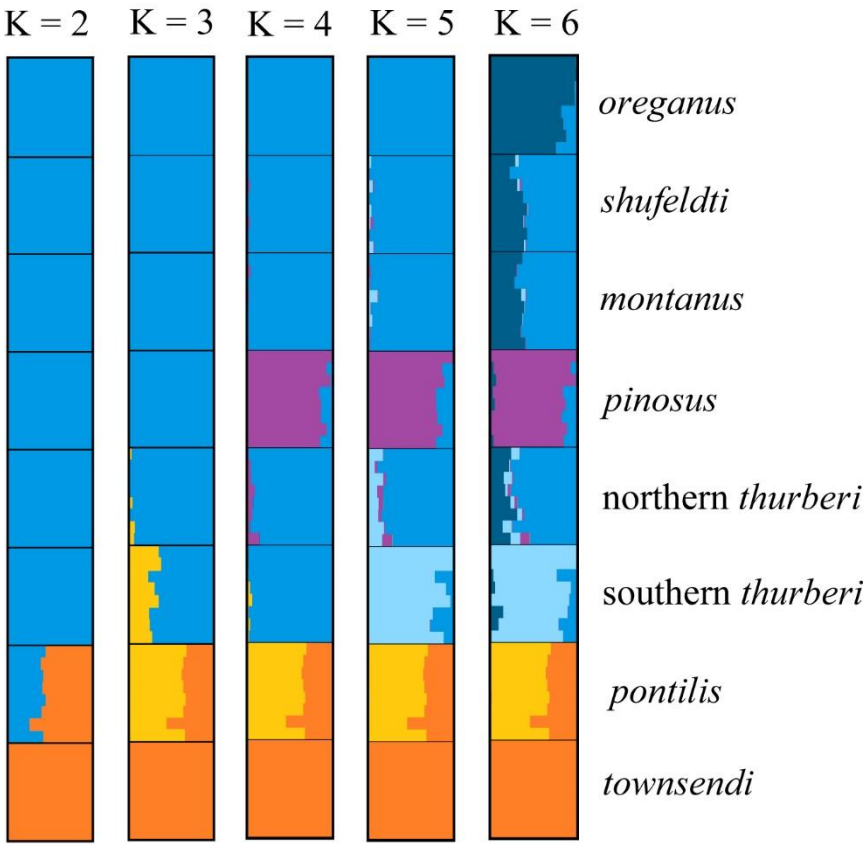
Results

Neutral genetic structure

The plot of the first two principal components from the PCA revealed four distinct clusters in the Oregon junco group. The most differentiated groups were *townsendi* and *pontilis* from Baja California, which formed two highly divergent clusters from each other and from other populations. A third, highly differentiated group corresponded to *pinosus* from coastal California, showing less differentiation than the Baja California forms with respect to a fourth cluster, which included all the remaining forms in the PCA (Fig. 3.1B). Within this fourth cluster, southern *thurberi* individuals presented certain degree of differentiation from the rest of the forms, a pattern more conspicuous when plotting the third and fourth components, which also revealed a slight signal of divergence in the *oreganus* form (Fig. 3.1C).

The STRUCTURE results were generally congruent with the PCA. The K = 2 plot recovered the *townsendi* as an independent population, which shared a considerable amount of variance with *pontilis*. In the analysis for K = 3, *pontilis* separated, congruently with the PCA, although southern *thurberi* showed some limited shared variance. The analysis for K = 4 identified the same four main clusters than the PC1 vs. PC2 plot. The plot for K = 5 clearly captured the differentiation of the southern *thurberi* form in a fifth cluster, with northern *thurberi* individuals showing a slight level of shared genetic variability. In the STRUCTURE plot for K=6, *oreganus* appears as an independent northern group with all individuals from northern *thurberi* and especially *montanus* and *shufeldti* forms presenting considerable probability of assignment to the *oreganus* cluster, increasing from south to north (Fig. 3.3).

Figure 3.3. Genetic structure of the Oregon junco forms based on 34,367 selectively neutral genome-wide SNPs using the program STRUCTURE. Each horizontal bar corresponds to an individual, with colors corresponding to posterior assignment probabilities to each of a number of genetic clusters (K). Colors correspond approximately to those in Fig. 3.1A.



Nei's distances and F_{ST} values were highly congruent with genetic structure analysis. The forms *townsendi*, *pontilis* and *pinosus* showed the highest values for both indices, while northern forms showed lower levels of differentiation, especially low for the southern against northern *thurberi* comparisons (Table 3.4).

Forward selection of explanatory variables

Out of eleven potentially relevant ecological variables for juncos, six were retained after the forward selection method intended for excluding non-significant effects. Retained variables included isothermality (BIO3), mean temperature of the warmest quarter (BIO10), mean precipitation of the warmest quarter (BIO18),

vegetation cover (TREE), greenness (NDVI), and elevation (SRTM) (Table 3.3). None of these variables was excluded due to excessive correlation as VIF values were below 10 (maximum recovered VIF = 5.75).

Redundancy analysis and variance partition

All six ecological variables retained in the forward selection method were included as explanatory variables in the RDA and the partial RDA models. RDA computes, in successive order, a series of axes that are linear combinations of the explanatory variables, and that best explain the variation in the matrix of response variables (Borcard et al. 2011). Six RDA axes (named RDA1 to RDA6 hereafter, ordered by the amount of variance explained by each one, reflected by the adjusted R^2) explained 6.26% of the total genetic variance in the non-conditioned model, and 1.18% when removing the effects of neutral genetic structure. The amount of explained variance increased to 36.61% when using only BayeScan outliers as response variables (Table 3.5). The permutation tests for the RDA models yielded a p-value below 0.001 in all three analyses, confirming the significance of the constraining variables effects.

Nei's\F_{ST}	*oreganus*	*shufeldti*	*montanus*	*pinosus*	N. *thurberi*	S. *thurberi*	*pontilis*	*townsendi*
oreganus		0.067	0.057	0.069	0.053	0.057	0.083	0.117
shufeldti	0.029		0.058	0.066	0.055	0.058	0.083	0.115
montanus	0.025	0.026		0.061	0.047	0.050	0.077	0.111
pinosus	0.032	0.032	0.028		0.058	0.060	0.087	0.119
N. *thurberi*	0.024	0.026	0.021	0.027		0.045	0.074	0.108
S. *thurberi*	0.027	0.028	0.023	0.028	0.021		0.074	0.109
pontilis	0.043	0.044	0.040	0.046	0.038	0.039		0.079
townsendi	0.061	0.061	0.057	0.063	0.056	0.057	0.043	

Table 3.4. Pairwise Nei's distances (lower diagonal) and F_{ST} values (upper diagonal) for all Oregon junco forms based on 9,436 independent, selectively neutral SNP loci.

Loadings of ecological explanatory variables on each one of the axes varied across the three different RDA models (Fig 3.4, Table3.5). In the non-conditioned RDA,

RDA1 had a large negative contribution of TREE and NDVI, and loaded positively on SRTM. The RDA2 loaded mostly on BIO3 and BIO10. The plot of per-individual projections on these two axes revealed a pattern generally similar to the PCA. The forms *townsendi*, *pontilis* and to a lesser extent *pinosus*, showed distinctive high values of correlation with both axes. The remaining forms showed similar correlation patterns with respect to RDA1, with southern *thurberi* individuals showing a clear association with RDA2 (Fig. 3.4A).

Table 3.5. RDA loads of the constraining variables in the first two axes and their explained variance for each one of the RDA models. The total variance explained by the full model (adjusted R^2 for the resultant six axes) is shown for each analysis. In all of the three analyses, the p-value for the full models was below 0.001. See Table 3.3 for variable definitions.

	Non conditioned RDA		**Partial RDA**		**RDA on BayeScan outliers**	
Adjusted R^2	6.26%		1.18%		36.61%	
Variable	RDA1	RDA2	RDA1	RDA2	RDA1	RDA2
BIO3	0.124	0.239	0.839	-0.427	0.214	0.763
BIO10	0.108	-0.610	0.359	-0.679	0.221	0.374
BIO18	0.052	0.021	-0.452	0.501	-0.031	-0.994
SRTM	0.646	-0.031	-0.324	-0.301	0.629	0.420
NDVI	-0.641	0.588	-0.320	0.199	-0.731	-0.277
TREE	-0.710	0.320	-0.270	0.326	-0.776	-0.423
Variance						
Eigenvalue	1304.262	567.705	481.320	350.714	63.655	12.657
Explained	0.025	0.011	0.003	0.002	0.250	0.050
Cumulative	0.025	0.037	0.003	0.005	0.250	0.300

In the partial RDA, the RDA1 axis had a large contribution of BIO3, while RDA2 loaded mostly on BIO10, and to a lesser extent BIO18 and TREE. Plotting these first two RDA axes revealed patterns of genetic correlation especially related to the first RDA axis for *pinosus*, which consequently presented the strongest association with

isothermality. The individuals from the southern population of *thurberi* reproduced the signal of association with RDA2 to a greater extent than in the non-conditioned RDA, evidencing again a positive correlation with the mean temperature of the warmest quarter. Northern forms of the Oregon junco (northern *thurberi*, *shufeldti*, *montanus* and *oreganus*) differentiated along the second RDA axis, with the northernmost *oreganus* showing the strongest association with the particularly conspicuous mean precipitation of the warmest quarter gradient, a pattern that was not visible in the non-conditioned RDA. The *townsendi* form from Baja California occupied positions closer to the origin of coordinates, suggesting a lower association between environmental and genetic variance (Fig. 3.4B).

The first axis of the RDA based on outlier loci showed moderate negative contributions from TREE and NDVI, and also a positive contribution from SRTM. Variance in BIO18 was almost entirely captured by RDA2, which also had a relatively high, negative contribution from BIO3. The plot showed a pattern of correlation between *pontilis* and *townsendi* along RDA1, while genetic variance in *oreganus* appeared strongly associated with the gradient of mean precipitation of the warmest quarter along RDA2. The rest of Oregon junco forms (and one atypical *oreganus* individual) were distributed in an opposite fashion, with small differences in their patterns of correlation with environmental variability captured in the second axis of the RDA (Fig. 3.4C).

The variance partition analysis showed that climate and neutral structure together explained 7.17% of the total genetic variability (Fractions A+B+C, Fig. 3.5). Since variable sets are not orthogonal, a 5.08% of variation was explained jointly by the environmental data and the first two components of the PCA based on neutral genetic positions (Fraction B, Fig.3.5). As recovered in the partial RDA, environmental variables alone explained 1.18% of the total variance (Fraction A, Fig. 3.5), while the non-overlapping fraction of neutral genetic structure explained 0.91% of the variability in the SNP dataset (Fraction C, Fig. 3.5). The p-value computed through the 10,000-step permutation test for each individual fraction was below 0.001 in all cases, thus confirming the significant effects of both variable sets.

Figure 3.4. Genetic-environment association analyses in the Oregon junco. Points represent the projection of individual genotypes on the first two RDA axes. Marker colors correspond to those on the range map on Fig. 3.1A. The explanatory variables are shown within the space defined by RDA1 and RDA2 by labeled vectors. Their contribution to each axis is represented by the length of their orthogonal projections over the blue scales on top and right sides of the graphs. Arrows indicate the direction of the gradient of variation for the corresponding environmental parameter. The value for each sample point on each explanatory variable can be obtained by an orthogonal projection on the corresponding plotted vector. (A) First two RDA axes of a non-conditioned RDA based on 11,261 SNPs. (B) First two RDA axes of a partial RDA based on 11,261 SNPs conditioned by neutral genetic structure, approximated by the first two PCs of a PCA based on neutral markers. (C) First two RDA axes of a non-conditioned RDA based on 87 SNP outliers identified in a BayeScan analysis.

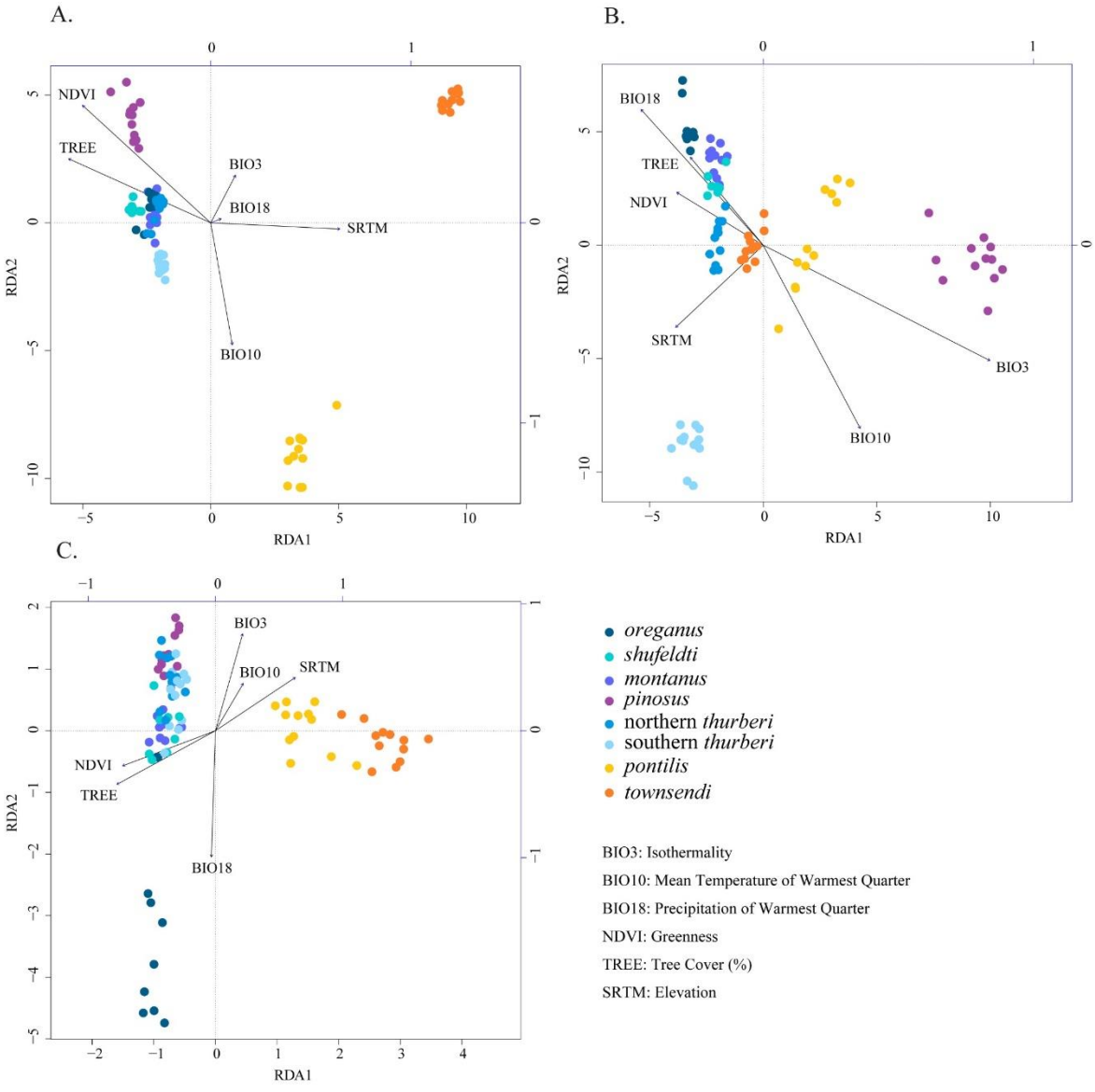
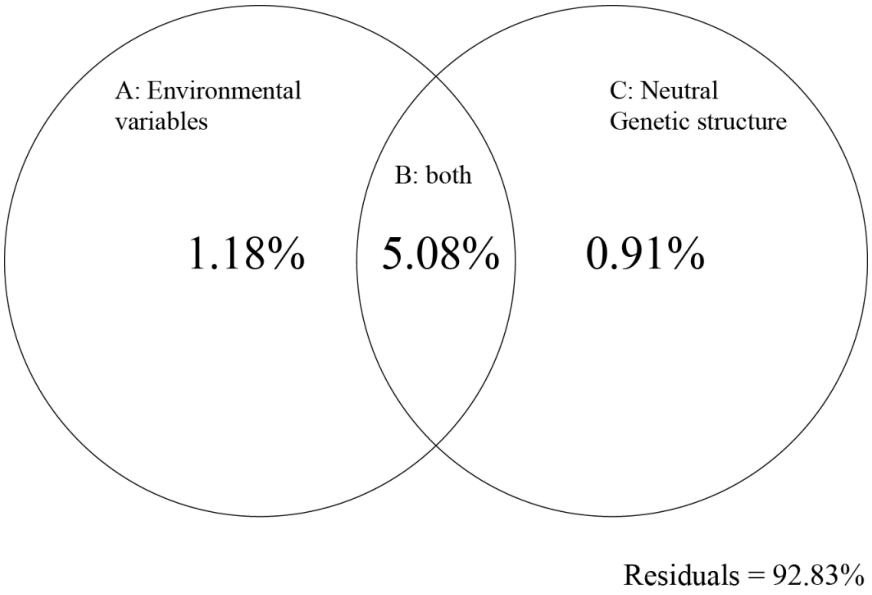


Figure 3.5. Plot of the fractions of genetic variability in Oregon juncos explained by (A) environmental variables alone; (B) the overlap of both environmental variables and neutral structure; (C) neutral genetic structure alone; and the unexplained genetic variability (residuals). P-values computed through a 1000-step permutation test for the fractions A, B and C, were below 0.001 in all cases.



Niche divergence tests

We tested for niche divergence and conservatism on each of the environmental variables. We found significant niche divergence between *pinosus* and northern *thurberi* for three of the six environmental variables analyzed (isothermality, precipitation of the warmest quarter and elevation; Table 3.6). When considering northern *thurberi* vs. southern *thurberi*, we found significant divergence for mean temperature of the warmest quarter and conservatism for isothermality (Table 3.6). NDVI was the only variable that exhibited significant divergence between *townsendi* and *thurberi* south (Table 3.6).

Genome scans

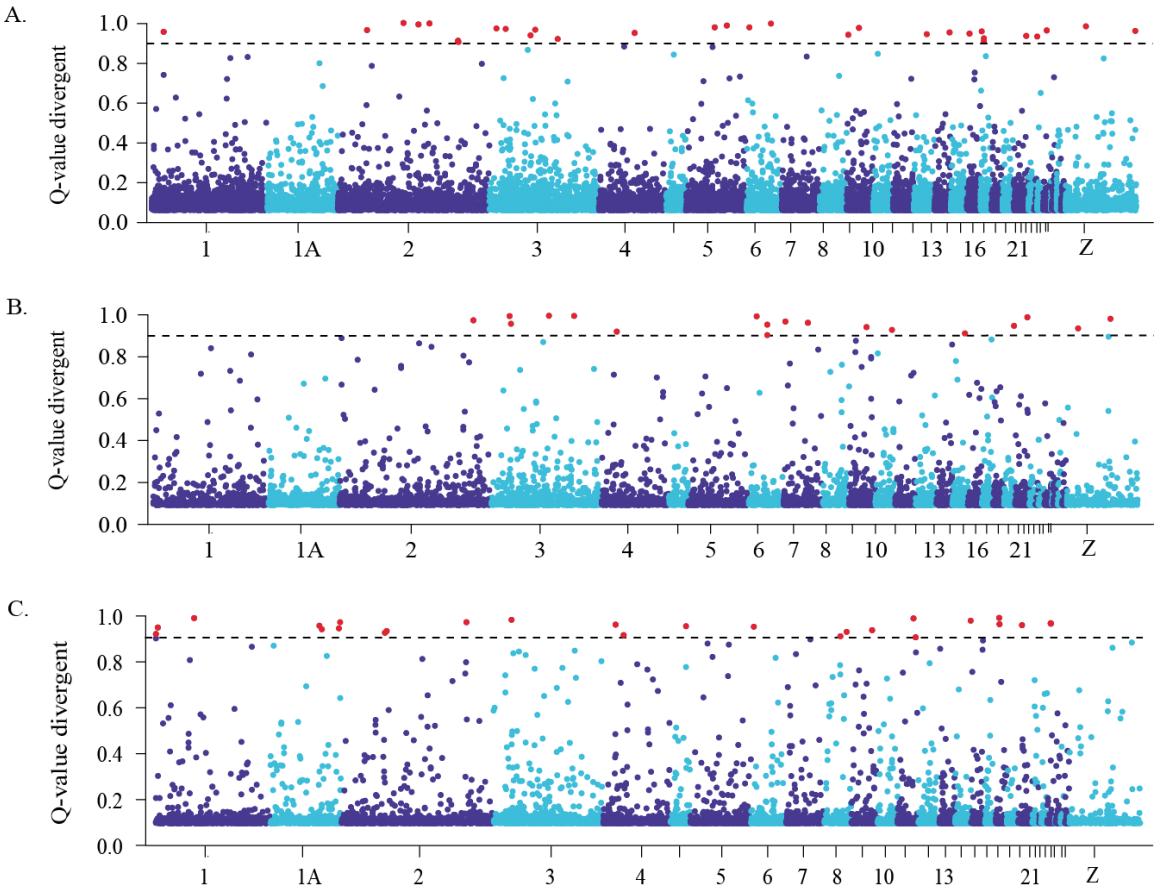
The BayeScan survey comparing all Oregon junco forms together detected 32 SNPs potentially under divergent selection, and 5 significant SNPs potentially under balancing selection. In the two pairwise comparisons *townsendi* vs. *pontilis* and

townsendi vs. *thurberi*, 20 and 30 significant SNPs with signs of having diverged under selection were detected, respectively, and no significant SNPs under balancing selection were found in either case. SNPs potentially differentiated under divergent selection appeared distributed across the genome in all comparisons, without obvious signs of heterogeneity among regions (Fig. 3.6). Chromosomes 1B and 16 harbored no SNPs so they are not shown in the plot.

Table 3.6. Results from the niche divergence test for *pinosus* vs. northern *thurberi*, northern vs. southern *thurberi* and southern *thurberi* vs. *townsendi*. Variables can show significant (p-value < 0.05) divergence ("Diverged") or conservatism ("Conserved") or show the expected divergence based on background **("**e.d.b.b.").

	Variable	**Result**	**p-value**	**Observed mean difference**	**Background mean difference**
pinosus vs. northern *thurberi*	TREE	e.d.b.b.	0.636	3.95	7.17
	BIO10	e.d.b.b.	0.718	20.40	17.34
	BIO3	**Diverged**	**0.000**	**16.78**	**9.45**
	BIO18	**Diverged**	**0.000**	**41.92**	**27.24**
	SRTM	**Diverged**	**0.032**	**1329.81**	**978.26**
	NDVI	e.d.b.b.	0.550	192.67	428.18
northern vs. southern *thurberi*	TREE	e.d.b.b.	0.580	2.11	4.52
	BIO10	**Diverged**	**0.030**	**41.85**	**23.87**
	BIO3	**Conserved**	**0.002**	**1.37**	**2.87**
	BIO18	e.d.b.b.	0.860	15.02	14.25
	SRTM	e.d.b.b.	0.474	202.60	140.29
	NDVI	e.d.b.b.	0.144	258.89	1019.97
southern *thurberi* vs. t*ownsendi*	TREE	e.d.b.b.	0.536	15.01	12.25
	BIO10	e.d.b.b.	0.412	47.88	43.03
	BIO3	e.d.b.b.	0.472	2.47	2.89
	BIO18	e.d.b.b.	0.188	74.77	67.26
	SRTM	e.d.b.b.	0.786	750.22	717.21
	NDVI	**Diverged**	**0.012**	**1597.62**	**574.28**

Figure 3.6. Plot of per-SNP posterior probability of divergence mediated by selection (shown as 1 – Q-value as computed by BayeScan) in (A) all Oregon junco forms together; (B) *townsendi* against *pontilis* and (C) *townsendi* against all *thurberi.* Loci above the dotted line (in red) are those below a false discovery rate of 10%.



Discussion

Neutral population structure and local adaptation explains genome-wide variance among Oregon junco forms

Our results reveal that both neutral and selective factors have played a role in driving divergence among Oregon junco populations, and that the relative contributions of geographic isolation and environment-driven selection are not uniform across the distribution range of the complex. The multivariate redundancy approach and variance partition found that environmental variables explained 1.18% of genomic variation when removing population structure variance; that

population structure explained 0.91% of genomic variation when removing variance that could be explained by environment; and that a relatively large 5.08% of genomic variation could be explained by both environment variation and population structure. Altogether, environment and neutral differentiation accounted for 7.17% of the variability in the 11,256 SNP matrix used as a set of response variables. Hence, residuals formed the remaining 92.83%, a fraction of the genomic variance non-structured in terms of neither environmental divergence nor population structure, potentially responding to loci under balancing selection or selective pressures not represented in our ecological variables, and to neutral variation shared by all Oregon junco forms because of their close relatedness and/or gene flow among them.

The amount of variance explained solely by environmental variables in our study was comparable to the values reported in studies applying RDA to detect specific correlations between genomic variation and a given set of potentially correlated variables, as shown in plants (e.g. Lasky et al. 2012; Vincent et al. 2013; De Kort et al. 2014) or other avian species (Safran et al. 2016; Szulkin et al. 2016). Previous studies on birds have used simple spatial variables such as geographic distance to control for the effects of spatial autocorrelation (Safran et al. 2016; Szulkin et al. 2016). Here, we controlled for genome-wide patterns of neutral variation by subtracting the variance captured by the first two PCs of a PCA based on neutral genome-wide SNPs, a method which should better account for population history and structure, including changes in effective population size, geographic isolation and related effects (Forester et al. 2016). Since spatial autocorrelation is usually accounted for by neutral genetic structure, we did not include spatial covariates to avoid over-conditioning the model (Rellstab et al. 2015). Given that only a small fraction of the surveyed genome is expected to be related to genes coding for climatic adaptation or linked to them (Meirmans 2015), a significant 1.18% of association between genomic variation and environmental variability in the conditioned (partial) RDA over only 11,261 genome-wide distributed SNPs is a compelling signal of local adaptation.

Genetic-environment association patterns in the diversification of the Oregon junco

The RDA revealed a number of strikingly different patterns of covariation between genetic variance and ecological variables likely to have played a role in Oregon junco diversification, especially when the effects of population history were removed. The forms *pontilis* and *townsendi* from Baja California, markedly isolated in terms of geography and neutral genetic variability, presented a low genetic-environment association when controlling for population history. This suggests that the differentiation between *townsendi* and *pontilis* is due largely to isolation in allopatry, in this case caused by unsuitable desert habitat in the lowlands surrounding their respective mountain ranges, a pattern also known as isolation by resistance (IBR, McRae and Beier 2007). In this scenario, our results suggest that differentiation is caused by drift under conditions of small population size and reduced gene flow due to geographic isolation, rather than divergent selection due to local adaptation, fitting the classic allopatric speciation model (Mayr 1942, 1963; Coyne and Orr 2004). This hypothesis is consistent with the niche divergence test comparing *townsendi* with southern *thurberi*, for which all tested variables but NDVI showed no signal of divergence beyond expectations based on background divergence.

The form *pinosus* showed considerable neutral genetic structure and a conspicuous pattern of genetic-environment association in both non-conditioned and partial RDA. When controlling for population structure, *pinosus* individuals showed high positive correlation values with isothermality (BIO3), while correlating negatively with elevation (SRTM). Indeed, *pinosus* presents the highest isothermality values, and the second lowest elevation after *oreganus*, reflecting a tolerance for low elevation conditions that are absent in neighboring *thurberi* (Miller 1941), a pattern also recovered in the niche divergence test. Unlike other differentiated forms like *pontilis* and *townsendi*, the *pinosus* form does not show high geographic isolation, and zones of intergradation with *thurberi* have been described (Miller 1941). Neutral population divergence despite the absence of geographic barriers to gene flow along with signs of local adaptation is a pattern consistent with isolation-by-adaptation, where barriers to gene flow may have arisen as individuals adapted to the distinct habitat of the coastal mountains of California. Niche distinctiveness and the genetic-environment association pattern of this form is thus congruent with a combination of warm latitude, low elevation and coastal influence that has

seemingly resulted in the adaptive differentiation of *pinosus* from the rest of the Oregon junco taxa. As a result, differentiation by drift may have led to positive correlations between adaptive and neutral genetic divergence (Nosil et al. 2008).

The southern *thurberi* individuals from Mont Laguna showed high overlap in terms of neutral genetic structure with northern *thurberi* and other boreal forms, and only slight differences in their genetic-environment association patterns when no controls for confounding factors were implemented. The Mount Laguna site represents the southernmost tip of the *thurberi* range in Southern California, which extends northward and reaches Oregon, forming a relatively continuous distribution (Miller 1941; Nolan et al. 2002), suggesting potentially high gene flow. However, the partial RDA revealed a distinct pattern of high correlation with the mean temperature of the warmest quarter (BIO10) for the southern *thurberi* juncos, differentiating them from the rest of Oregon forms. They also correlated negatively with the mean precipitation of the warmest quarter (BIO18). This pattern seems congruent with the habitat of Mount Laguna, and in general with the southern inland range of Oregon juncos, quite arid during summer but subject to snowfall in winter due to the high elevations (Miller 1941), contrasting sharply with the more climatically moderate coastal and northern populations. The limited neutral genetic structure between *thurberi* range extremes but considerable differentiation in the genetic-environment association patterns is consistent with a process of local adaptation across a selective gradient (Forester et al. 2016), in which selection is the prominent evolutionary force driving differentiation (Haldane 1948; Slatkin 1973; Nagylaki 1975; Felsenstein 1976) while neutral alleles may move freely across space. However, the niche divergence test comparing southern and northern *thurberi* populations only recovered a signal of significant divergence in mean temperature of the warmest quarter (BIO10). Furthermore, we found statistical support for niche conservation in isothermality (BIO3), a variable for which the RDA recovered certain difference in the pattern of association between the two populations. This suggests that either the niche divergence test failed in detecting actual patterns of ecological differentiation, perhaps due to low sample size or small effect size; or that the RDA overestimated the correlation between genetic and most of the environmental variables, perhaps because of population structure related

effects that the treatment failed to correct for. Nevertheless, the signal of differentiation due to local adaptation in terms of temperature was consistent in both analyses.

The boreal Oregon junco forms including *oreganus*, *montanus*, *shufeldti*, and *thurberi* individuals from Tahoe, presented a more conspicuous pattern of local adaptation along a shallow selective gradient. These forms showed very low neutral genetic structure or differences in ecological covariances in the non-conditioned RDA, yet showed an increasing signal of association following their latitudinal distribution in the partial RDA. A strong association pattern emerged especially for mean precipitation of the warmest quarter (BIO18) and for correlated environmental variables of tree coverage (TREE) and greenness (NDVI), matching quite precisely their latitudinal distribution along a gradient of increasing humidity and vegetation cover. The ecology-related differences in genetic variance, consistent with the taxonomic classification of these forms, is especially relevant considering the relative phenotypic similarity of these taxa, and their apparent intergradation (Miller 1941; Nolan et al. 2002).

GEA methods present a number of limitations, including potentially high rates of type I error. Here, rather than detecting specific loci under selection, we aimed to explore how selection and neutral processes shape the variability in Oregon juncos, but the risk of finding false significant associations between genetic variance and ecological parameters persists. To further test the environmental associations revealed in this study, we implemented a highly conservative approach using only BayeScan outliers as response variables in the model. The non-conditioned RDA based on 87 SNP loci identified by BayeScan as potential targets of divergent selection yielded relatively lower resolution than the partial RDA. BayeScan has been shown to produce relatively few false positives, but it is also a conservative approach, the sensitivity of which decreases with selection strength (Narum and Hess 2011). The RDA suggests that BayeScan correctly identified outliers related to low temperatures and high precipitation for *oreganus* samples, a pattern congruent with the habitat and with the outcomes of previous analyses for this form. It also detected highly differentiated positions in *pontilis* and *townsendi* that correlate with

RDA1, but in this case, associations with specific environmental variables were lower, and the pattern disappears in the RDA based on the entire SNP dataset when correcting for population structure. This may suggest that BayeScan failed to exclude the effects of demographic history, or in turn, that controlling for the genetic variance captured by the PCA was overly conservative. BayeScan was also less successful in detecting adaptive divergence in *pinosus,* and especially in northern Oregon junco forms, where selection may be weaker or have acted during a shorter period. Nevertheless, the outlier SNP dataset explained 36.61% of the total climatic variability, a considerable amount compared with the full SNP data RDA models, indicating a good fit of the retained outliers to the linear regression on the environmental parameters.

Interactions among environment, geography and demographic processes results in three different modes of divergence within the Oregon junco lineage

The Oregon junco is one of the six phenotypically and genetically differentiated dark-eyed junco taxa evolved during a northward expansion from Central America after the last glacial maximum, c.a. 18,000 years ago (Milá et al. 2007; Chapter I, Friis et al. 2016). Similar postglacial expansions have been reported for many other bird species (Seutin et al. 1995; Milá et al. 2000; 2006; Hansson et al. 2008; Malpica and Ornelas 2014; Alvarez et al. 2015). However, the population structure documented in this study reveals a variety of different spatial, selective and demographic factors not previously documented in other avian taxa. In light of our genetic-environment association analyses, the patterns recovered by the PCA and STRUCTURE analysis reveal at least three different effects of geography and demographic history interacting to varying degrees with selection in the process of Oregon junco diversification. First, the IBR pattern of differentiation presented by *pontilis* and *townsendi* may suggest that these forms are peripheral remnants of an original, broader distribution of the Oregon juncos, thereafter isolated in the 'sky islands' of Baja California and diverging predominantly by drift. Indeed, in his thorough monograph on the geographic variation in juncos, Alden Miller (1941) had perceptively suggested early on that the habitat of Oregon juncos from Baja California did not seem to account for their phenotypic differentiation from Californian forms, and that their distinctive traits appeared to be predominantly

"historical" (p. 306). The spatial configuration and recovered patterns of neutral and adaptive divergence for *pontilis* and *townsendi* fit the first model of neutral divergence of isolated population in approximately similar habitats proposed in the Introduction. Second, the IBA pattern found in *pinosus* suggests that the current area of intergradation corresponds to a secondary contact zone that formed after diverging in relative isolation, maybe linked to the ancient coastal closed-cone pine forest that has allegedly diminished since the Pleistocene (Miller 1941), a scenario that differs from that of historically continuous boreal forms, and from the Baja California forms, which seemingly never came into secondary contact with any other Oregon junco form. The second mode of divergence between parapatric populations proposed in the Introduction has been only partially confirmed, since local adaptation seems to have resulted in reduced levels of gene flow, leading to increased neutral genome-wide differentiation (Nosil et al. 2008; Funk et al. 2011; Flaxman et al. 2014). Third, the geographic continuum represented by *thurberi*, *shufeldti*, *montanus* and *oreganus* is also captured in the STRUCTURE analysis by a gradual signal of differentiation following a latitudinal distribution, suggesting that ongoing gene flow may occur among forms. Combined with the signal of increasing environmental association recovered in the partial RDA, these outcomes are highly consistent with a process of differentiation driven by local adaptation along a selective gradient in the direction of the northward expansion, fulfilling the third hypothesized mode of divergence for these forms of Oregon junco. Interestingly, ecological association approaches have been shown to perform better along clines of selection where demographic expansions align with the gradient of ecological variables, usually related to latitude (Frichot et al. 2015), as is the case in the *Junco* system.

A relevant aspect of the marked population structure found among Oregon junco forms is that it is based on a relatively small subset of genome-wide SNPs randomly sampled from across the genome, representing a genomic fraction not greater than 0.2% of the total of 1.2 Gb. The clear signal of divergence mediated by environmental factors recovered also in the RDAs indicates that divergence may have taken place at the level of the entire genome, suggesting the role of multiple selective pressures consistent with local adaptation along the latitudinally broad and heterogeneous

distribution of the Oregon juncos. The presence of outliers potentially under positive selection scattered across the genome seems to support this hypothesis of selection-driven genome-wide divergence, rather than widespread drift among isolated populations. Other examples of such patterns of genomic differentiation due to divergent selection at early (e.g. Parchman et al. 2013; Brawand et al. 2014; Egan et al. 2015) and intermediate (e.g. Riesch et al. 2017) stages of speciation have been reported in the recent literature, contrasting with proposed models of speciation initiated by divergent selection in a few, localized genes involved in reproductive isolation (e.g. Nadeau et al. 2012; Poelstra et al. 2014).

Conclusion

Our analyses reveal the role of both local adaptation and demography in driving rapid diversification of the Oregon junco during its northward colonization of western North America since the last glacial maximum. The combined effects of a demographic expansion along a selective gradient with a heterogeneous landscape of environmental variability have resulted in a striking array of divergence modes within a single lineage, from isolated forms in Baja California that have differentiated largely by drift in isolated 'sky islands', to adaptive diversification along selective gradients with no obvious geographic barriers to gene flow across *thurberi* and the northernmost forms. There is also a compelling example of isolation by adaptation in the case of *pinosus*, where ecological barriers to gene flow seem to maintain its divergence with respect to nearby forms. Genome-wide patterns of divergence indicate that Oregon junco diversification has been driven by multiple ecological factors acting on many loci across the genome, and suggests that selection may promote local adaptation in short periods of time, highlighting the role of adaptive divergence in the early stages of the speciation process. Future analyses of dense sequencing and functional gene characterization will be necessary to further identify adaptive changes promoting barriers to gene flow and reveal the genomic architecture of rapid diversification.

References

Alvarez, S., J. F. Salter, J. E. McCormack, and B. Milá. 2015. Speciation in mountain refugia: phylogeography and demographic history of the pine siskin and black-capped siskin complex. Journal of Avian Biology 47:335-345.

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222:1810-1819.

Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology 20:2044-2072.

Billiard, S., T. Lenormand, and S. Gavrilets. 2005. Evolution of migration under kin selection and local adaptation. Evolution 59:13-23.

Blanchet, F. G., P. Legendre, and D. Borcard. 2008. Forward selection of explanatory variables. Ecology 89:2623-2632.

Borcard, D., F. Gillet, and P. Legendre. 2011. Chapter 6. Canonical Ordination. Pp. 153-226. Numerical Ecology with R. Springer.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, and E. Bezault. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375.

Carson, H. L. 1975. The genetics of speciation at the diploid level. The American Naturalist 109:83-92.

Chhatre, V. and K. Emerson. 2016. StrAuto: Automation and parallelization of STRUCTURE analysis. See http://strauto. popgen. org.

Christmas, M. J., E. Biffin, M. F. Breed, and A. J. Lowe. 2016. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. Molecular Ecology 25:4216-4233.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Darwin, C. 1859. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

De Kort, H., K. Vandepitte, H. H. Bruun, D. Closset-Kopp, O. Honnay, and J. Mergeay. 2014. Landscape genomics and a common garden trial reveal adaptive

differentiation to temperature across Europe in the tree species Alnus glutinosa. Molecular ecology 23:4709-4721.

Dray, S., P. Legendre, and G. Blanchet. 2009. packfor: Forward Selection with permutation. R package version 0.0-7/r58.

Dwight, J. 1918. The geographical distribution of color and of other variable characters in the genus *Junco*: a new aspect of specific and subspecific values. Bull. Am Mus. Nat. Hist. 38:269-309.

Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. Proceedings of the National Academy of Sciences of the United States of America 101:975-979.

Egan, S. P., G. J. Ragland, L. Assour, T. H. Powell, G. R. Hood, S. Emrich, P. Nosil, and J. L. Feder. 2015. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. Ecology letters 18:817-825.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of applied ecology 41:263-274.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103:285-298.

Excoffier, L. and N. Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends in ecology & evolution 23:347-351.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological Speciation: an integrative approach. Molecular ecology 23:513-521.

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. Annual review of genetics 10:253-280.

Flaxman, S. M., A. C. Wacholder, J. L. Feder, and P. Nosil. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. Molecular ecology 23:4074-4088.

Foll, M. and O. E. Gaggiotti. 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977-993.

Forester, B. R., M. R. Jones, S. Joost, E. L. Landguth, and J. R. Lasky. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular ecology 25:104-120.

Frichot, E., S. D. Schoville, P. de Villemereuil, O. E. Gaggiotti, and O. François. 2015. Detecting adaptive evolution based on association with ecological gradients: orientation matters! Heredity 115:22-28.

Friis, G., P. Aleixandre, R. Rodríguez-Estrella, A. G. Navarro-Sigüenza, and B. Milá. 2016. Rapid postglacial diversification and long-term stasis within the songbird genus Junco: phylogeographic and phylogenomic evidence. Molecular Ecology 25:6175-6195.

Funk, D. J., S. P. Egan, and P. Nosil. 2011. Isolation by adaptation in Neochlamisus leaf beetles: host-related selection promotes neutral genomic divergence. Molecular Ecology 20:4671-4682.

Goudet, J., T. Jombart, and M. J. Goudet. 2015. Package 'hierfstat'.

Haldane, J. B. 1948. The theory of a cline. Journal of genetics 48:277-284.

Hansson, B., D. Hasselquist, M. Tarka, P. Zehtindjiev, and S. Bensch. 2008. Postglacial colonisation patterns and the role of isolation and expansion in driving diversification in a passerine bird. PLoS One 3:e2794.

Hedrick, P. W., M. E. Ginevan, and E. P. Ewing. 1976. Genetic polymorphism in heterogeneous environments. Annual review of Ecology and Systematics 7:1-32.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Jones, M. R., B. R. Forester, A. I. Teufel, R. V. Adams, D. N. Anstett, B. A. Goodrich, E. L. Landguth, S. Joost, and S. Manel. 2013. Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. Evolution 67:3455-3468.

Kawecki, T. J. and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology letters 7:1225-1241.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources 15:1179-1191.

Lasky, J. R., D. L. Des Marais, J. McKAY, J. H. Richards, T. E. Juenger, and T. H. Keitt. 2012. Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Molecular Ecology 21:5512-5529.

Legendre, P. and L. Legendre. 1998. Numerical ecology: second English edition. Developments in environmental modelling 20.

Lischer, H. E. L. and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Malpica, A. and J. F. Ornelas. 2014. Postglacial northward expansion and genetic differentiation between migratory and sedentary populations of the broad-tailed hummingbird (*Selasphorus platycercus*). Molecular Ecology 23:435-452.

Manthey, J. D. and R. G. Moyle. 2015. Isolation by environment in white-breasted nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. Molecular Ecology 24:3628-3638.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

Mayr, E. 1947. Ecological factors in speciation. Evolution:263-288.

Mayr, E. 1954. Change of genetic environment and evolution.

Mayr, E. 1963. Animal species and evolution. Belknap Press, Cambridge, MA.

McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens, and R. T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66:526-538.

McCormack, J. E., A. J. Zellmer, and L. L. Knowles. 2010. Does Niche Divergence Accompany Allopatric Divergence in Aphelocoma Jays as Predicted Under Ecological Speciation?: Insights from Tests with Niche Models. Evolution 64:1231-1244.

McRae, B. H. and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences 104:19885-19890.

Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to avoid them. Molecular Ecology 24:3223-3231.

Milá, B., P. Aleixandre, S. Alvarez-Nordström, J. McCormack, E. Ketterson, and J. Atwell. 2016. More than meets the eye: Lineage diversity and evolutionary history of Dark-eyed and Yellow-eyed juncos. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco (ED Ketterson and JW Atwell, Editors). University of Chicago Press, Chicago, Illinois, USA:179-198.

Milá, B., D. J. Girman, M. Kimura, and T. B. Smith. 2000. Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proceedings of the Royal Society Biological Sciences Series B 267:1033-1040.

Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society B-Biological Sciences 274:2653-2660.

Milá, B., T. B. Smith, and R. K. Wayne. 2006. Postglacial population expansion drives the evolution of long-distance migration in a songbird. Evolution 60:2403-2409.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Mitton, J., Y. Linhart, J. Hamrick, and J. Beckman. 1977. Observations on the genetic structure and mating system of ponderosa pine in the Colorado Front Range. Theoretical and Applied Genetics 51:5-13.

Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. ffrench-Constant, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences 367:343-353.

Nadeau, S., P. G. Meirmans, S. N. Aitken, K. Ritland, and N. Isabel. 2016. The challenge of separating signatures of local adaptation from those of

isolation by distance and colonization history: The case of two white pines. Ecology and evolution 6:8649-8664.

Nagylaki, T. 1975. Conditions for the existence of clines. Genetics 80:595-615.

Narum, S. R. and J. E. Hess. 2011. Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources 11:184-194.

Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus, S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Nosil, P. 2012. Ecological Speciation. Oxford University Press.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:"isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336.

Ojeda Alayon, D. I., C. K. Tsui, N. Feau, A. Capron, B. Dhillon, Y. Zhang, S. Massoumi Alamouti, C. K. Boone, A. L. Carroll, and J. E. Cooke. 2017. Genetic and genomic evidence of niche partitioning and adaptive radiation in mountain pine beetle fungal symbionts. Molecular Ecology 26:2077-2091.

Oksanen, J., F. Blanchet, R. Kindt, P. Legendre, and R. O'Hara. 2016. Vegan: community ecology package. R Packag. 2.3-5.

Parchman, T., Z. Gompert, M. Braun, R. Brumfield, D. McDonald, J. Uy, G. Zhang, E. Jarvis, B. Schlinger, and C. Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology 22:3304-3317.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS genet 2:e190.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Pritchard, J. K. and A. Di Rienzo. 2010. Adaptation–not by sweeps alone. Nature Reviews Genetics 11:665-667.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

R_Core_Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R_Studio_Team. 2015. RStudio: Integrated Development for R. R Studio, Inc., Boston, MA.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24:4348-4370.

Riesch, R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K. Lucek, E. Hellen, V. Soria-Carrasco, and S. R. Dennis. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. Nature Ecology & Evolution 1:0082.

Rundle, H. and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336-352.

Safran, R., E. Scordato, M. Wilkins, J. K. Hubbard, B. Jenkins, T. Albrecht, S. Flaxman, H. Karaardıç, Y. Vortman, and A. Lotem. 2016. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. Molecular Ecology 25:3865-3883.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Seutin, G., L. M. Ratcliffe, and P. T. Boag. 1995. Mitochondrial DNA homogeneity in the phenotypically diverse redpoll finch complex (Aves: Carduelinae: *Carduelis flammea-hornemanni*). Evolution 49:962-973.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733-756.

Soberon, J. and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas.

Sork, V. L., K. Squire, P. F. Gugger, S. E. Steele, E. D. Levy, and A. J. Eckert. 2016. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, Quercus lobata. American journal of botany 103:33-46.

Szulkin, M., P. A. Gagnaire, N. Bierne, and A. Charmantier. 2016. Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. Molecular ecology 25:542-558.

Templeton, A. R. 1981. Mechanisms of speciation-a population genetic approach. Annual review of Ecology and Systematics 12:23-48.

Termignoni-García, F., J. P. Jaramillo-Correa, J. Chablé-Santos, M. Liu, A. J. Shultz, S. V. Edwards, and P. Escalante Pliego. 2017. Genomic footprints of adaptation in a cooperatively breeding tropical bird across a vegetation gradient. Molecular Ecology.

Turner, S. D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv:005165.

Uyeda, J. C., S. J. Arnold, P. A. Hohenlohe, and L. S. Mead. 2009. Drift promotes speciation by sexual selection. Evolution 63:583-594.

Van Den Wollenberg, A. L. 1977. Redundancy analysis an alternative for canonical correlation analysis. Psychometrika 42:207-219.

Vincent, B., M. Dionne, M. P. Kent, S. Lien, and L. Bernatchez. 2013. Landscape genomics in Atlantic salmon (Salmo salar): searching for gene–environment interactions driving local adaptation. Evolution 67:3469-3487.

Wang, I. J. and G. S. Bradburd. 2014. Isolation by environment. Molecular Ecology 23:5649-5662.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. Evolution 62:2868-2883.

Wright, S. 1943. Isolation by distance. Genetics. 28:114-138.

Wright, S. 1946. Isolation by distance under diverse systems of mating. Genetics 31:39.

Yates, A., W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, and L. Gil. 2016. Ensembl 2016. Nucleic acids research:gkv1157.

Zheng, X. 2012. SNPRelate: parrallel computing toolset for genome-wide association studies. R package version 95.

CHAPTER IV: Genome-wide evidence for a demographic expansion and selection in the rapid postglacial radiation of the juncos

Junco bairdi, picture by Borja Milá

Guillermo Friis, Ángeles de Cara, Borja Milá

Abstract

Rapid evolutionary radiations are often the result of the combined effects of selective pressures and demographic processes. The reconstruction of the evolutionary history of the genus *Junco* revealed a set of recently diversified lineages that originated during the postglacial expansion and recolonization of the North American continent, contrasting with more divergent and geographically isolated ancestral southern forms. Signs of geographic and historical factors, along with specific analyses evidencing the contribution of natural and sexual selection in the radiation of the northern forms of *Junco* suggest that the process of lineage diversification was driven by multiple evolutionary forces. Here we combine whole-genome and genotyping-by-sequencing (GBS) data from four different forms of junco to (i) test the demographic expansion of the juncos using genome-wide data, (ii) explore the importance of gene flow at different stages of the radiation, and (iii) identify genomic regions potentially under selection that may have played a role in lineage divergence. We use Tajima's D and G-PhoCS (generalized phylogenetic coalescent sampler) to test the population-expansion and recent-divergence hypotheses in northern junco forms. We also perform per-locus genomic surveys of F_{ST} and D_{XY} to infer selection-mediated divergence during the postglacial radiation. Both Tajima's D and G-PhoCS revealed recent demographic expansions for all the northern junco forms included in the analysis, reinforcing the hypothesis of multiple lineage differentiation driven by a postglacial northward recolonization of North America. In addition, G-PhoCS revealed limited gene flow among young lineages of junco in the diversification process, suggesting that early divergence took place in allopatry. Per-locus genome scans found no specific regions of high differentiation but rather a number of highly divergent variants scattered across the genome, suggesting the role of multiple selective pressures acting on numerous loci from the early stages of the speciation process.

Key words:

Demographic expansion, postglacial recolonization, gene flow, recent diversification, lineage divergence, selection.

Introduction

Lineage diversification processes during the colonization of newly available habitats generally result from the combined effects of divergent selection and changes in the effective population size as new populations become established and distribution ranges expand (Simpson 1953; Schluter 2000). In recently diversified systems, signs of divergent selection at the genomic level can be more easily detected, so inferences about how adaptation drives lineage diversification and speciation can be made (Nosil et al. 2008; Feder et al. 2012; Riesch et al. 2017). However, strong demographic changes such as bottlenecks and rapid population expansions interact with selective forces by affecting parameters such as standing genetic variation or the rates of allele fixation (Mayr 1942, 1963). Differences in rates of genetic recombination across the genome and gene flow throughout the diversification process and between contemporary populations are also crucial in shaping the genomic patterns of divergence and lineage diversification (Nosil 2008; Via 2009; Feder et al. 2012; Li et al. 2012; Burri et al. 2015). Thus, both demographic history and potential selective forces need to be taken into account when studying adaptive radiations (Jensen et al. 2007; Li et al. 2012; Laurent et al. 2016).

An example of recent radiation during the recolonization of newly available habitats is provided by the young lineages of the dark-eyed junco *Junco hyemalis* complex, a case of rapid diversification across a latitudinal gradient during the northward expansion from Central to North America as ice-sheets retreated after the last glacial maximum (LGM). Previous analyses based on mtDNA (see Chapter I, Friis et al. 2016) confirmed the postglacial origin of the young northern lineages ca. 18,000 years ago, while phylogeographic analyses recovered clear signs of population expansion of dark-eyed junco forms across North America as evidenced by star-like haplotype networks and significant negative values of Fu's F_S (Fu 1997). Analyses of neutral genetic structure and phylogenomic reconstructions based on genome-wide SNP markers revealed considerable population divergence and resolved the polytomic relationships of specific and subspecific forms of yellow-eyed and dark-eyed juncos. The marked patterns of genetic differentiation are consistent with a process of fast lineage diversification among populations with limited gene flow, possibly as populations became established in high-elevation sky islands across the

patchy distribution of suitable habitat found across the junco range (Miller 1941; Nolan et al. 2002). Along with geographic and historic factors related to demographic expansion and divergence in isolation, we found signs of sexual selection and local adaptation involved in lineage diversification (see Chapters II and III, respectively), revealing a complex scenario where the combined effects of neutral and selective evolutionary forces produced one of the fastest cases of speciation known in vertebrates. However, analyses based on genome-wide data are still needed to test the role of potential bottlenecks and posterior expansions and to assess the levels of gene flow during the junco radiation.

Here we revisit the outcomes and test the proposed hypothesis of Chapter I with an array of genomic data and analytical tools. Using seven fully sequenced genomes from individuals belonging to four distinct junco taxa representative of both old and young lineages, we apply coalescence analysis to simultaneously test the recent origin of the North American forms of junco, infer the demographic history of both highly differentiated and recently radiated forms, and estimate the role of gene flow among lineages. We also use genotyping-by-sequencing (GBS, Elshire et al. 2011) markers to compute overall values of Tajima's neutrality test (Tajima 1989) and conduct genomic surveys based on pairwise comparisons of per-locus values of F_{ST} (Weir and Cockerham 1984) and D_{XY} (Nei 1987) genetic distances/statistics, to further study patterns of genomic divergence and demographic evolution in the *Junco* complex.

Materials and Methods

Whole-genome alignment and variant calling

We conducted the variant calling of seven genomes corresponding to the forms *bairdi* (n = 2, sampled in Sierra de la Laguna, Baja California, Mexico), *phaeonotus* (n = 2, sampled in Tascate, Chihuahua, Mexico), *caniceps* (n = 2, sampled in the Toiyabe Mountains, Nevada, USA) and *thurberi* (n = 1, sampled in Mount Laguna, California, USA) (Table 4.1). We excluded from these analyses the higher-quality eighth genome to avoid bias due to differences in depth coverage (see General Methods).

For the specific analyses of the present chapter, we used Trim Galore (Krueger 2015) instead Trimmomatic (Bolger et al. 2014) for read quality filtering (see General Methods). We removed Illumina adapters, trimmed low-quality ends of the reads to a genotyping phred quality score of 20, and filtered out any reads with a final length below 40. Duplicated reads were removed with PICARDTOOLS v1.126 (http://broadinstitute.github.io/picard/). Remaining reads were then mapped against the reference junco genome using the mem algorithm in the Burrows-Wheeler Aligner (BWA, Li and Durbin 2009). We used the Genome Analysis Toolkit (GATK, McKenna et al. 2010) version 3.6-0 to realign reads around indels using the IndelRealigner tool and then we applied the HaplotypeCaller tool to call the individual genotypes. Finally, we used the GenotypeGVCFs tool to gather all the per-sample GVCF files generated in the previous step, with the option *-allsites* to produce a set of jointly-called positions keeping both variant and invariant sites (GATK Best Practices, DePristo et al. 2011; Auwera et al. 2013) in the variant call format (*vcf*).

Table 4.1. Junco forms and number of genotyped individuals per locality included in the study. State abbreviations are the following: Nevada (NV), Utah (UT), Colorado (CO), California (CA) and Arizona (AZ) in the USA; Durango (DGO), Mexico City (CM), Chihuahua (CHIH) and Baja California (BC) in Mexico.

Form	*Localities (State)*	*Whole Genome*	*GBS*
caniceps	Toiyabe Range (NV); Uinta Mountains, Wasatch Mountains, Timpanogos, Dixie N.F. (UT); Aspen, Rico, Sangre de Cristo Mountains (CO).	2	16
oreganus	Mount Laguna (CA).	1	16
phaeonotus	Pinaleno Mountains (AZ), Bajío de la Víbora (DGO), Tascate (CHIH), La Cima (CM).	2	16
bairdi	Sierra de la Laguna (BC).	2	8
Total		7	56

Genotyping-by-sequencing

For the set of analyses based on GBS SNPs reported in this chapter, we used the individual genotypes from 56 juncos belonging to the same taxa represented by the whole genomes, corresponding to (sample sizes in parentheses): *bairdi* (8), *phaeonotus* (16), *caniceps* (16) and *oreganus* (16) (Table 4.1).

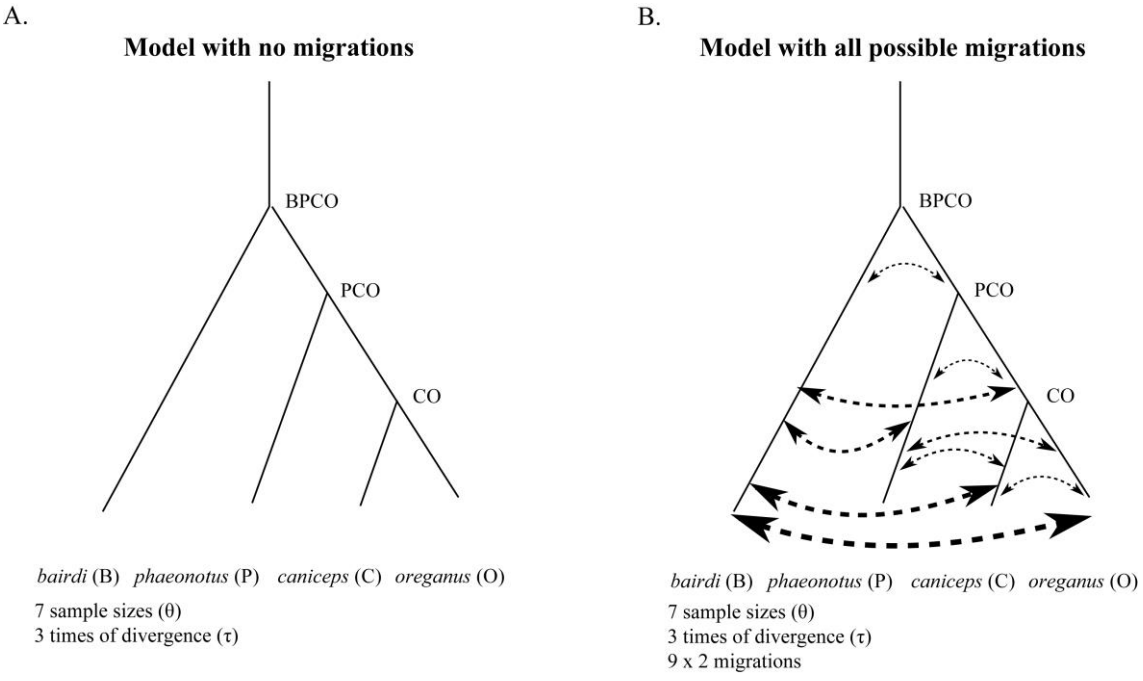
Demographic inferences

We inferred divergence times, ancestral population sizes and rates of post-divergence gene flow using G-PhoCS (Gronau et al. 2011) and the 7 fully sequenced genomes. After testing for convergence with different prior distributions, we used exponential distributions with a mean of 0.001 for divergence times, a mean of 0.01 for effective population sizes and Gamma distributions for gene flow rates with alpha=0.002 and beta=0.00001. The analysis was conditioned on the phylogenetic relationships based on mtDNA and SNP markers inferred in the first Chapter for *bairdi*, *phaeonotus*, *caniceps* and *oreganus.* We kept the data mapped to those scaffolds with size above that N50, and selected loci of 1kb separated by 40kb in order to avoid linked regions. We obtained ca. 9000 loci in which we evaluated the convergence of Markov chains of 500,000 iterations. Each chain was evaluated independently for convergence, and we used seven that converged for posterior analyses. In each chain, we removed the first 100,000 iterations to obtain the posterior distributions for each parameter inferred.

To rescale coalescent units to years and number of individuals, we calibrated the divergence times using the time to the most recent ancestor (TMRC) inferred for the node between *bairdi* and the remaining forms of junco, based on a molecular clock for mtDNA markers and estimated at 0.4M years (see Chapter I; Friis et al. 2016). Using the relationship $\mu = \tau_{div} / T_{DIV}$ (where μ is the mutation rate; τ_{div} is the divergence time in coalescent units, and T_{DIV} is the divergence time in years for the split of the ancestral population), and assuming a generation time of 1 year, mutation rate was thus obtained by $\mu = \tau_{div} / T_{DIV}$, while effective population size was obtained by rescaling θ following $Ne = \theta / (4\mu)$. We measured gene flow using the total migration rate during the length of a band (Gronau et al. 2011; Freedman et al. 2014), that is, the inferred migration rate multiplied by the time during which

migration took place. We followed the same lax criteria as Freedman et al. (2014) to consider gene flow as significant: either the 95% highest posterior density (HPD) interval of the total migration rate did not include 0, or the estimated total migration rate was above 0.03 with a posterior probability greater than 50%. We performed independent analyses using two different scenarios of migration between divergent lineages, a first one without migration, and a second one that included all possible migration bands. A sketch of these models is shown in Fig.4.1.

Figure 4.1. Graphic representation of the demographic models tested and the parameters inferred, assuming no migration (A) and assuming migration among lineages (B). Phylogenetic relationships among taxa are based on the phylogenetic reconstructions from Chapter 1 (Friis et al. 2016). Divergence times (τ parameter) are estimated for each node (denoted by the combined initials of the taxa included in the corresponding clade) with the exemption of the B-PCO, used for scaling (see text). Sample sizes (θ parameter) are estimated for each internal and external branch. Arrows represent bidirectional potential migrations to be tested.



Tajima's D boxplots

Complementarily to G-PhoCS analyses, we calculated Tajima's D (Tajima 1989) to test for an excess of rare variants, consistent with recent population grow. For that

purpose, we retained the eight samples of each form with a lower proportion of missing sites for a final number of 32 samples. We constructed a data matrix of biallelic SNPs excluding those out of a range of coverage between 2 and 100 or with less than the 75% of the individuals genotyped. Because Tajima's D relies on the level of both low and high frequency polymorphisms, we did not apply a minor allele frequency filter for this analysis. Instead, we increased the genotyping phred quality score threshold to 90 to retain only high-confidence variants. We did implement a threshold for SNPs showing highly significant deviations from Hardy-Weinberg equilibrium (HWE) with a p-value of 10^{-4} to filter out false variants arisen by the alignment of paralogous loci. We use DnaSP v6 (non published Beta version, previous DnaSP v5: Librado and Rozas 2009) to compute various genetic indices. In order to define real, fully sequenced loci over which to compute Tajima's D, F_{ST} and D_{XY} statistics (see below), we edited the *vcf* files by means of customized R scripts (see Annex IV for R code): Taking advantage of the fixed length of our GBS reads (69 pair of bases), we modified the *vcf* file so in the CHROM field each 69 pb loci was labeled as a single chromosome or scaffold. We then fed the *vcf* files to DnaSP v6 for computing Tajima's D over each of the resulting loci. Boxplots for each of the forms were created with the boxplot2 function in the R-package gplots (Warnes et al. 2013).

Genome scans

Highly differentiated loci that appear as outliers in an F_{ST} distribution can be interpreted as potential targets of divergent selection standing out in a background of balanced or neutrally maintained genomic variation (Faria et al. 2014; Rellstab et al. 2015). However, F_{ST}-like indices are relative measures of divergence and thus highly dependent on within-population levels of variation (Cruickshank and Hahn 2014). Therefore, signs of high F_{ST} in specific regions of the genome may be due not to strong divergent selection but to low within-population diversity. On the contrary, D_{XY} (average number of nucleotide substitutions per site between two given populations, Nei 1987), is an absolute measure of differentiation among populations (Nei and Li 1979; Cruickshank and Hahn 2014). Here we performed pairwise genome scans of the different forms of junco using D_{XY} and F_{ST} (Weir and Cockerham 1984) over non-overlapping windows of 69 bp, corresponding to fully

sequenced GBS reads. For a fair trade-off between recovery of confidently mapped GBS reads and the length of surveyed genomic regions, we mapped the reads against the white-throated sparrow (*Zonotrichia albicollis*) genome [GenBank accession: ARWJ00000000] (Romanov et al. 2011), assembled in 6,018 scaffolds with a N50 of 4.9 Mb and a L50 of 52 (e.g. Campagna et al. 2017). The white-throated sparrow is more closely related to juncos than the zebra finch (whose genome was used as a reference in the genome scans of Chapter III) so the mapping yields a higher number of confidently aligned reads; at the same time, the available genome is assembled in considerably longer scaffolds than our junco reference genome, allowing to inspect patterns of genomic divergence across longer regions. We applied the same set of tools and parameters than for junco genome mapping and genotyping for GBS samples described in the General Methods section. We conducted the comparative analysis for *phaeonotus* (16 individuals) against a set of eight individuals of *caniceps* and 8 individuals of *oreganus* grouped into a single population; for *oreganus* against *caniceps* (16 individuals of each form); and for *bairdi* against *caniceps* (8 individuals of each form). For each one of these analyses we built SNP matrices retaining only biallelic SNPs with coverage between 2 and 100 and a genotyping phred quality score over 70. Positions with less than 25% of the individuals genotyped were removed from each data matrix, along with those presenting a MAF below 0.09 when comparing *bairdi* against *caniceps*, and 0.04 in the rest. Once again, we implemented a p-value threshold for HWE of 10^{-4} to filter out false variants caused by the alignment of paralogous loci (see Table 4.2 for final datasets sizes). To compute D_{XY} and F_{ST} over the fully sequenced loci we applied the same treatment as for the computation of Tajima's D, and we fed the modified *vcf* file to DnaSP v6 to obtain the average number of nucleotide differences between two given populations, known as K_{XY}, along with the per-locus F_{ST} values. To compute D_{XY}, we finally divided the K_{XY} index by 69, the length of the GBS reads. Genome scan plots were conducted in R 3.2.2 (R Core Team 2015, see Annex IV for R code) using the package qqman (Turner 2014). For the sake of clarity in the graphics, we only plotted those variants found in scaffolds larger than 2 Mb, spanning 73% of the complete genome. To set a minimum threshold and detect outliers potentially under selection, we applied the R function boxplots.stats and implemented a restrictive

coefficient of 3, *i.e.* a value 3 times higher than the length of the third and fourth interquartile range for the D_{XY} and F_{ST} distributions.

Table 4.2. Sample sizes per taxon and number of SNP loci obtained. General filters included a depth range of from 2 to 100 and a p-value threshold for Hardy-Weinberg deviation of 0.0001.

Analysis	**Samples**	**QUAL**	**MAF**	**SNPs (loci)**	**Missing data**
Tajima's D:					
oreganus	8	90	None	61,410 (39,983)	25%
caniceps	8	90	None	48,284 (32,000)	25%
phaeonotus	8	90	None	42,710 (29,151)	25%
bairdi	8	90	None	16,032 (12,964)	25%
Genome Scans:					
bairdi vs. *caniceps*	16	70	0.09	10,114 (8,347)	75%
phaeonotus vs. *caniceps* and *oreganus*	32	70	0.04	15,356 (11,749)	75%
caniceps vs. *oreganus*	32	70	0.04	20,380 (15,207)	75%

Results

G-PhoCS analyses

Each run of G-PhoCS of 500,000 samples took between 3 days using the multithread version for the model with 10 variables and no migration, and over 12 days without multithreading for the model with all possible migration bands. The model including no migration bands recovered contemporary effective population sizes showing high demographic expansions for the three young junco lineages here analyzed (*phaeonotus*, *caniceps* and *oreganus*), with increases ranging by factors of approximately 3 to 10 times the size of the immediate ancestral populations. Conversely, the divergent *bairdi* form showed a population reduction by a factor of nearly 5 (Fig. 4.2A, Table 4.3). The model with migration recovered less pronounced

values of demographic change for young northern lineages (by factors of 1, 2 and 2.5 ca. for *oreganus*, *phaeonotus* and *caniceps*, respectively), yet revealed a strong population size reduction for *bairdi* by a factor of ~170 (Fig. 4.2B, Table 4.3).

Dating of lineage splits varied markedly between the models with and without migration when fixing the B-PCO divergence at 400,000 years (where 'B', 'P', 'C' and 'O' refer to *bairdi*, *phaeonotus*, *caniceps* and *oreganus* respectively, and combined initials denote corresponding ancestral populations from here on). Estimated divergence times (T_{DIV} P-CO and T_{DIV} C-O, where T_{DIV} means time of divergence) showed little difference within each model, yet were nearly eight times higher in the model without migration. In the latter, they ranged from 370K to 393K years, while in the model with migration they ranged from 44K to 56K years (Table 4.3).

Table 4.3. Effective population sizes (N_e) in number of individuals, divergence times (T_{DIV}) in years, and total migration rates (indicated by arrows) estimated by G-PhoCS for the models with and without migration, along with the 95% confidence intervals of the higher posterior density range (HPD). Internodal, ancestral populations and divergence times are denoted by the combined initials of the taxa included in the corresponding clade (see Fig. 4.1). Significant migration rates are marked with asterisks.

Model	Mean	Low 95% HPD	High 95% HPD
No migration model			
N_e BPCO	1,542,461	1,510,153	1,575,384
N_e *bairdi*	333,846	324,615	342,769
N_e PCO	866,153	318,153	1,419,384
N_e *phaeonotus*	2,237,230	2,158,153	2,318,461
N_e CO	208,000	14,153	467,384
N_e *caniceps*	2,061,230	1,986,153	2,137,230
N_e *oreganus*	760,615	728,307	792,615
T_{DIV} BPCO	400,000	391,692	408,000
T_{DIV} PCO	384,000	374,461	392,923
T_{DIV} CO	379,384	369,538	388,923

Table 43. (Continuation).

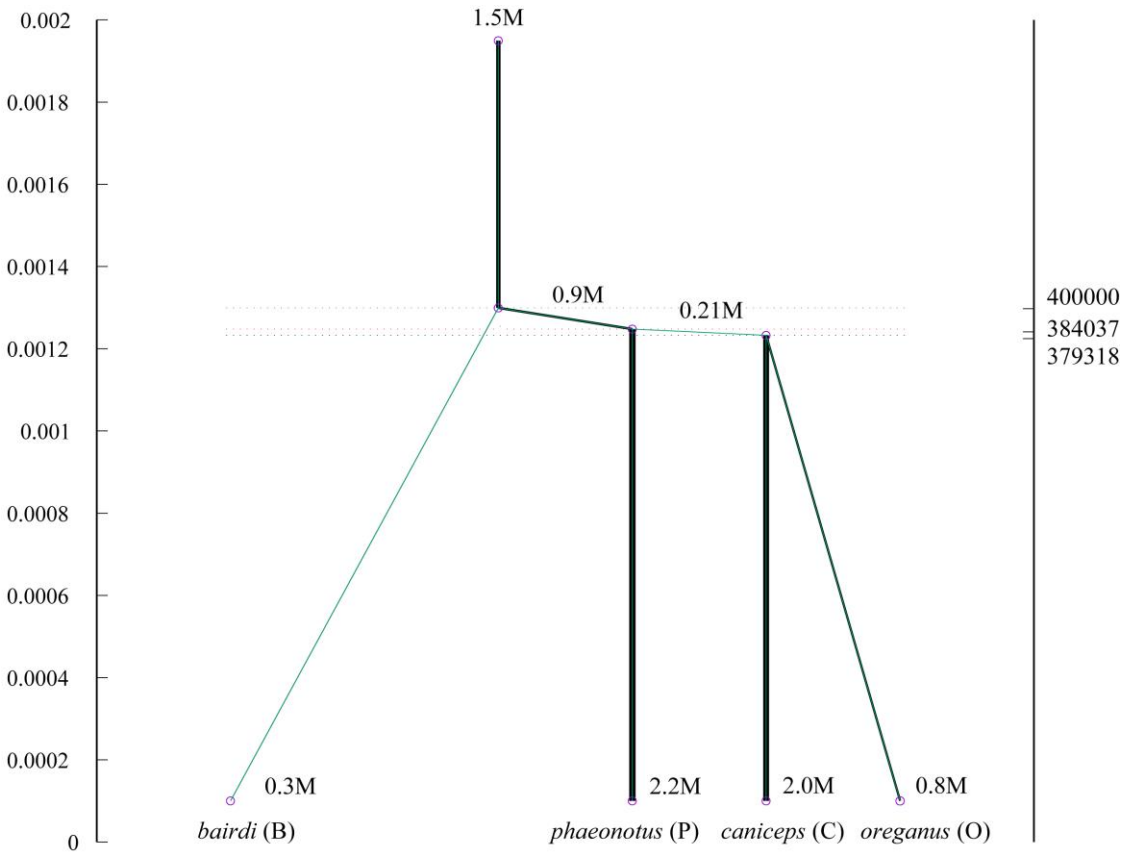
Model	Mean	Low 95% HPD	High 95% HPD
Migration model			
N_e BPCO	4,635,616	3,298,772	6,039,341
N_e *bairdi*	26,538	25,344	27,702
N_e PCO	130,039	124,003	135,107
N_e *phaeonotus*	237,090	220,436	251,385
N_e CO	84,913	35,196	145,327
N_e *caniceps*	217,290	206,798	227,605
N_e *oreganus*	82,270	77,050	87,355
T_{DIV} BPCO	400,000	378,014	423,028
T_{DIV} PCO	50,342	47,234	56,416
T_{DIV} CO	44,446	40,386	47,213
bairdi -> PCO	0.004	0.000	0.017
PCO -> *bairdi**	4.431	3.789	5.046
phaeonotus -> CO	0.002	0.000	0.009
CO -> *phaeonotus*	0.329	0.000	1.038
bairdi -> CO*	0.086	0.065	0.108
CO -> *bairdi**	1.433	1.239	1.640
bairdi -> *phaeonotus*	0.001	0.000	0.005
phaeonotus -> *bairdi**	0.063	0.044	0.084
phaeonotus -> *oreganus*	0.004	0.000	0.020
oreganus -> *phaeonotus**	0.036	0.022	0.050
phaeonotus -> *caniceps*	0.023	0.000	0.054
caniceps -> *phaeonotus**	0.061	0.022	0.100
bairdi -> *caniceps*	0.000	0.000	0.001
caniceps -> *bairdi**	0.084	0.062	0.106
caniceps -> *oreganus*	0.020	0.000	0.068
oreganus -> *caniceps*	0.023	0.000	0.048
bairdi -> *oreganus*	0.001	0.000	0.007
oreganus -> *bairdi**	0.040	0.026	0.055

The analysis modeling migration revealed diverse signals of gene flow among all current and ancestral lineages of junco included in the test. Following criteria in Freedman et al. (2014), significant although limited gene flow was detected between *oreganus* and *phaeonotus*, with a migration rate of 3.6%; and between *caniceps* and

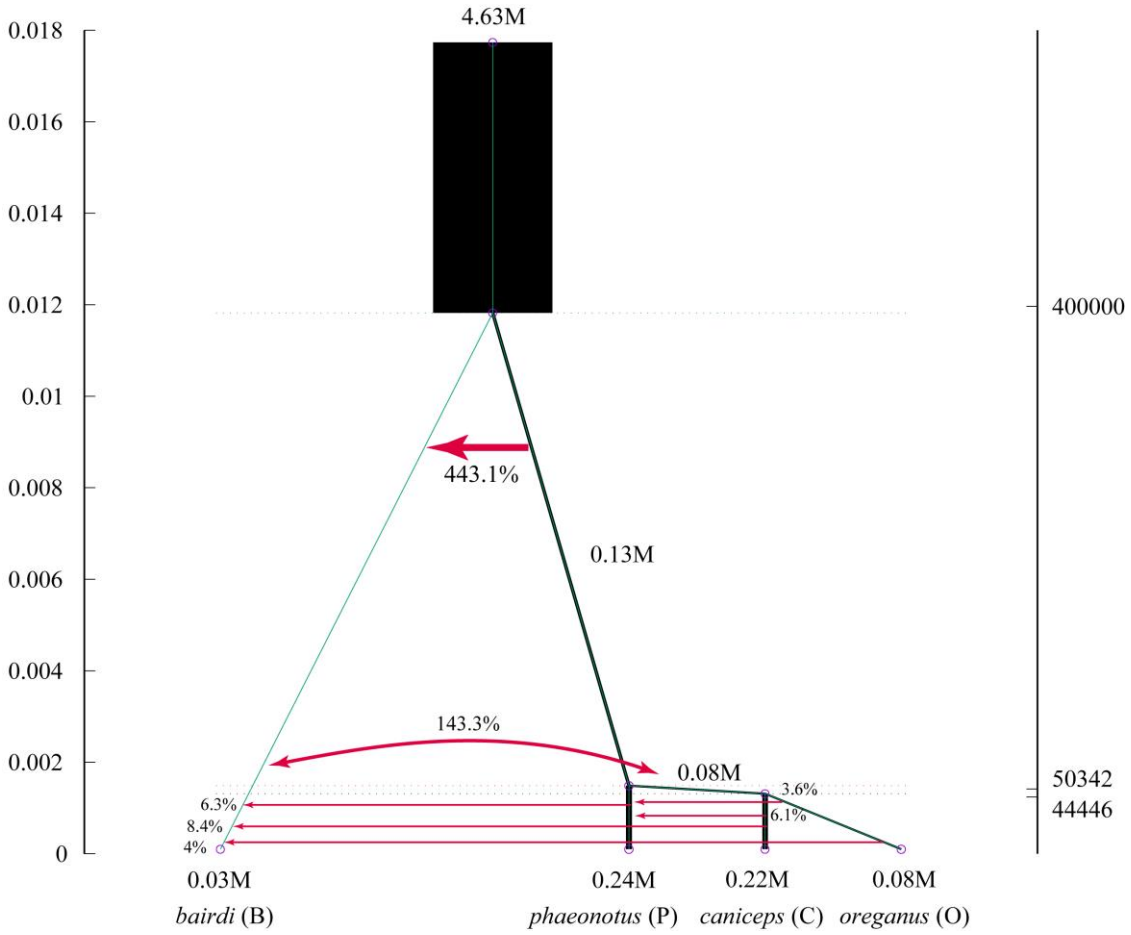
phaeonotus with a rate of 6.1%. High migration rate of 443.1% was detected between PCO and *bairdi*. The analyses also recovered a high migration rate of a 143.3% from the ancestral CO population and *bairdi*, and a reverse rate *bairdi* -> CO of 8.6%. *Bairdi* also presented signs of incoming gene flow from *phaeonotus* (6.3%), *caniceps* (8.4%) and *oreganus* (4%) (Table 4.3).

Figure 4.2. Evidence for a recent demographic expansion of the young northern junco forms. Divergence times in coalescent units (left vertical axis) and years (right vertical axis), effective population sizes (in millions of individuals, represented by branch thickness), and significant rates of gene flow (in percentage, represented by red arrows) inferred by G-PhoCS in a model with no migrations(A), and a model incorporating all possible migrations (B). (C) Boxplots of genomewide Tajima's D values per taxon.

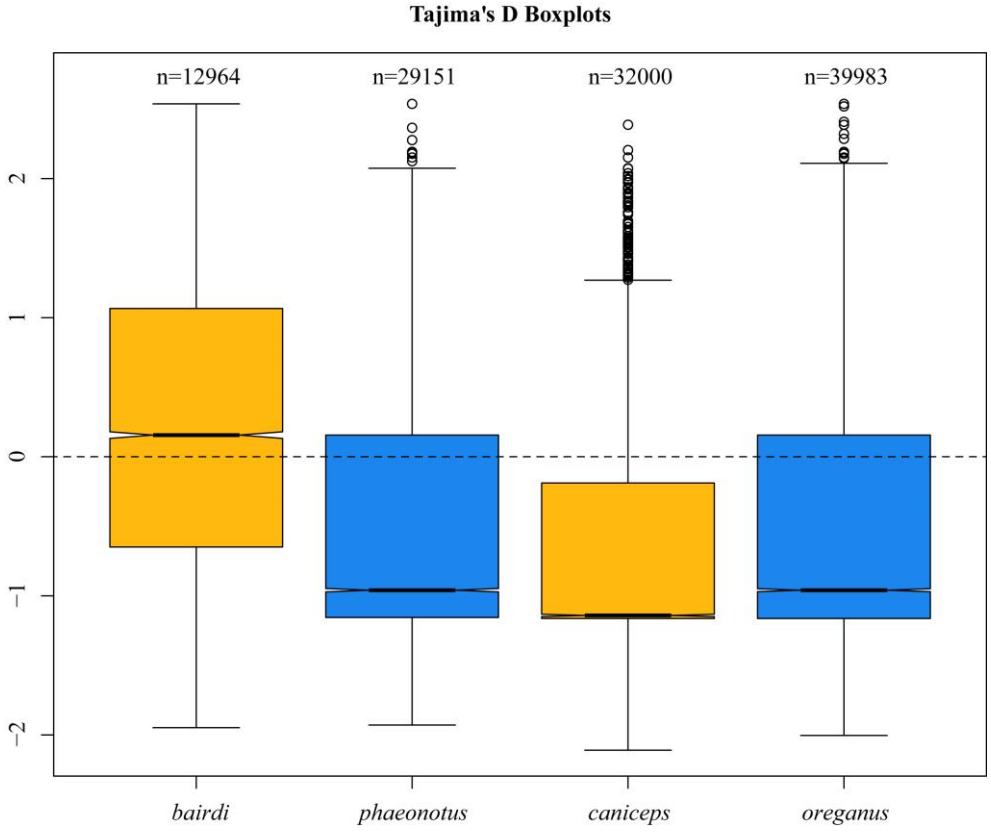
A.



B.


4.63M
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
400000
50342
44446
443.1%
0.13M
143.3%
0.08M
3.6%
6.1%
6.3%
8.4%
4%
0.03M
0.24M
0.22M
0.08M
bairdi (B)
phaeonotus (P)
caniceps (C)
oreganus (O)

C.


Tajima's D Boxplots
n=12964
n=29151
n=32000
n=39983
2
1
0
−1
−2
bairdi
phaeonotus
caniceps
oreganus

Tajima's D

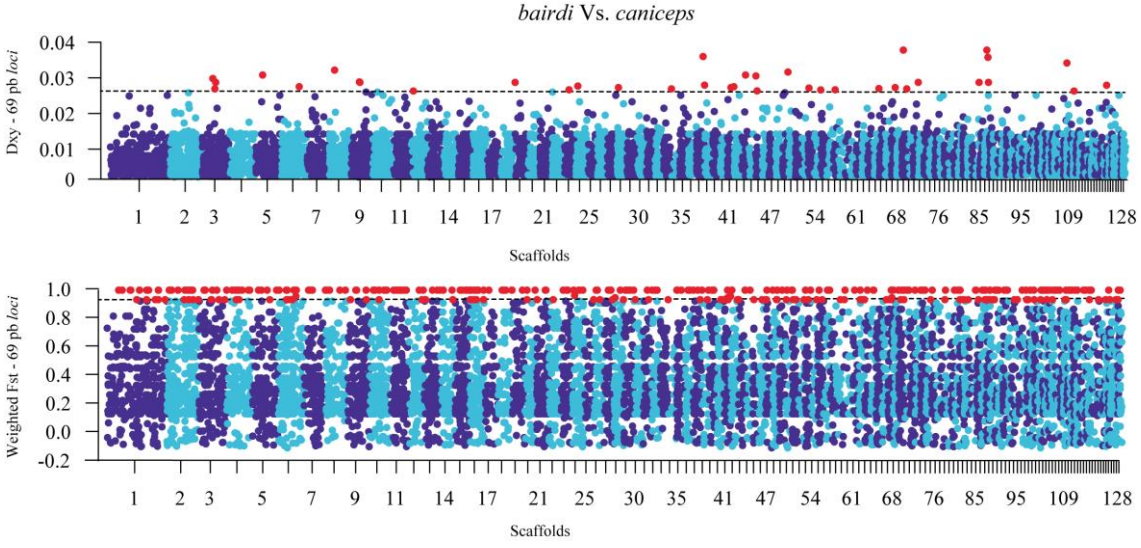
Tajima's D scores were congruent with G-PhoCS outcomes. The values were consistently negative for the young lineages *phaeonotus*, *caniceps* and *oreganus*, revealing an excess of rare variants that is consistent with a recent demographic increase. In contrast, *bairdi* showed a majority of loci with a positive Tajima's D score, suggesting a reduction of effective sample size (Fig. 4.2C).

Genome scans

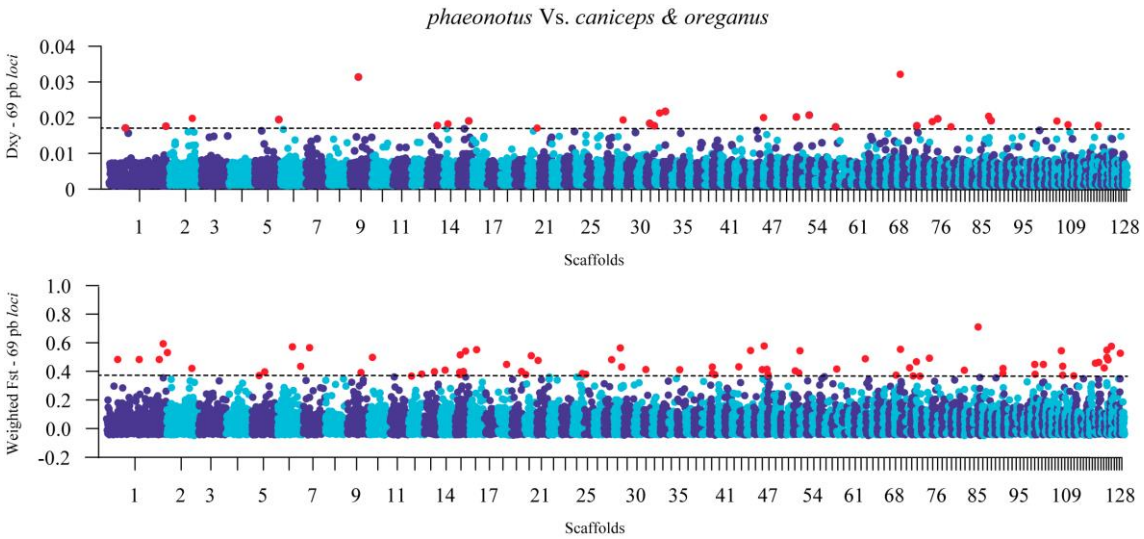
Both D_{XY} and F_{ST} revealed a pattern of low heterogeneity in the profile of differentiation across the genome, with no clear regions of relatively higher divergence among compared taxa. Overall values of divergence were also congruent with the phylogenetic relationships of the compared taxa. The *bairdi* versus *caniceps* comparison (Fig. 4.3A) presented an elevated number of fixed positions and a corresponding pattern of high D_{XY} scores, while *phaeonotus* and dark-eyed form comparisons yielded lower divergence values, especially in the *caniceps* versus *oreganus* analysis (Fig. 4.3B and 4.3C).

Figure 4.3. Genomic patterns of divergence based on pairwise F_{ST} and D_{XY} comparisons based on GBS data for (A) *bairdi* against *caniceps* (8 samples of each form); (B) *phaeonotus* (16 samples) against a set of eight samples of *caniceps* and 8 samples of *oreganus* grouped into a single population; and (C) for *oreganus* against *caniceps* (16 samples of each form). Outliers are marked in red.

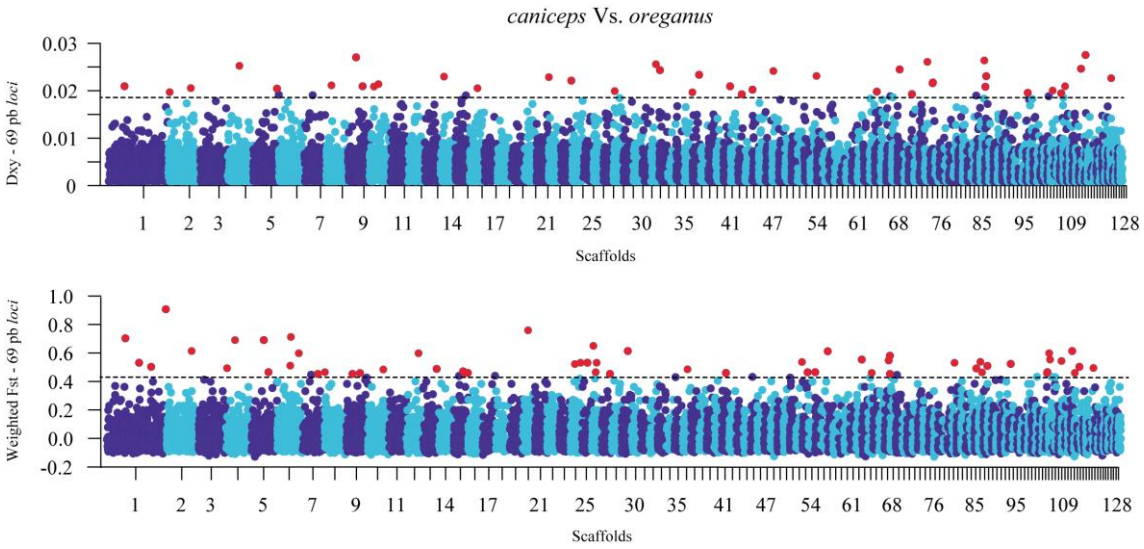
A.



B.



C.



Discussion

Signs of demographic expansion during the diversification of the Junco *complex*

Our analyses recovered a general pattern of considerable decreases in effective population sizes followed by strong demographic expansions for the young junco lineages, and a marked signal of population reduction in the older *bairdi* lineage from the tip of Baja California after its cladogenesis. In relative values, similar patterns of demographic evolution were recovered in the models with and without migration. However, absolute population estimates of extant lineages and especially

of times of divergence in the model without migration showed low congruence with general geographic ranges and previous molecular dating of recent lineages, one order of magnitude above the times estimated by BEAST and IMa2 (see Chapter I, Friis et al. 2016). In contrast, the model incorporating all possible migration bands yielded more consistent divergence times and population sizes, with the exemption of the demographic size of the ancestral BPCO population, which seems clearly overestimated in the model including gene flow. Nevertheless, because incorporating gene flow to the model depicts a more realistic scenario during the diversification of multiple clades, we place higher confidence in the estimates recovered in the model with migration.

The pattern of demographic growth after population reduction events recovered in the G-PhoCS analyses were especially clear for *phaeonotus* and *caniceps* regardless of migration. The strong bottleneck signal of *bairdi*, consistent with the restricted and isolated geographic distribution of the form, was more pronounced in the model with migration. Fu's tests (Fu 1997) from Chapter I based on mtDNA markers recovered a marked signal of population expansion for *phaeonotus*, but failed to detect the demographic growth for *caniceps*, or the bottleneck of *bairdi*. Conversely, Fu's F_S yielded a clear signal of expansion for the *oreganus* form, while the G-PhoCS mean estimates of effective population size for the ancestral CO population and the *oreganus* current lineage were very similar.

The disagreement between the results of the migration model implemented in G-PhoCS and Fu's F_S about the demographic history of *oreganus* may be due primarily to the fact that G-PhoCS assumes a constant population size for all branches of the phylogeny on which the analysis is conditioned, preventing the detection of initially small populations that grow at a later time (Gronau et al. 2011), a likely scenario for a founder *oreganus* group undergoing a range expansion, as well as for the remaining recent lineages of junco. More severe bottlenecks in all recent lineages were possibly unnoticed likewise. In addition, the lack of resolution when estimating population sizes along such short branches of the phylogeny may be precluding the detection of demographic changes from the CO to the *oreganus* lineage, as suggested by the wide 95% HPD interval for the ancestral CO population

size, which overlaps completely with the corresponding *oreganus* interval. Other possible explanations are that the *oreganus* individual used for the analysis was sampled in Mount Laguna, a relatively small population presenting signs of geographic isolation (see Chapter III) that may not reflect the footprint of overall demographic expansion. Alternatively, the fact that the *oreganus* taxon is represented by only one genome in the G-PhoCS analysis may have introduced a bias in the model inferences.

Tajima's D analyses also supported the bottleneck for *bairdi* and the demographic expansion for the recently diversified lineages, yielding a highly consistent pattern with the hypothesis of a postglacial demographic expansion in the northern juncos. A potential bias may result from the substantially lower number of loci for the *bairdi* analyses (Clark et al. 2005), caused by the availability of higher quality sequenced samples for the rest of the forms, although this seems unlikely considering the high number of nearly 13,000 representative loci included in the *bairdi* analysis.

Coalescent analyses based on genome-wide data support the recent origin of the northern juncos

Divergence times for the phenotypically differentiated, young northern forms of junco estimated by G-PhoCS ranged between 47.2K and 56.4K years for the P-CO split, and between 40.4K and 47.2K years for the C-O split in the model with migrations. These inferences fall within the TMRCA estimated by BEAST (see Chapter I), and reflect a recent origin for both the yellow-eyed *phaeonotus* and the two dark-eyed *caniceps* and *oreganus* forms. The congruence between G-PhoCS and BEAST dating also reflects considerable consistency in the estimated rates of divergence for genome-wide SNPs and mtDNA markers. Taking into account isotopic geological dating of the LGM (Clark et al. 2009) and the gathered evidence a about the evolutionary history of the genus, a slightly earlier origin for the northern juncos coinciding with appearance of suitable habitats in the current distribution seems likely.

Limited gene flow and selection-driven genome-wide divergence during the recent radiation of the juncos

The G-PhoCS analysis recovered a general pattern of relatively low migration rates among closely related young forms of *Junco*, contrasting with extremely high signs of gene flow among ancestral lineages. Following criteria in Freedman et al. (2014), significant but limited gene flow among recent lineages were detected for *oreganus* to *phaeonotus*, and for *caniceps* to *phaeonotus*. We also recovered a sensitively higher migration rate for the ancestral CO lineage to *phaeonotus* (32.9%), but the 95% HPD interval including zero prevents its consideration. Overall, these outcomes are congruent with the hypothesized scenario of rapid differentiation with limited gene flow proposed in previous chapters of this dissertation, based on the recurrent patterns of neutral genetic structure among northern junco lineages obtained in our analyses. An alternative explanation is that G-PhoCS lacks sufficient power to detect gene flow over periods as short as the time elapsed between the young lineages of junco separated and the present time, resulting in an underestimation of gene flow rates.

In sharp contrast, an extremely high migration rate was detected for the PCO -> *bairdi* band, indicating high levels of gene flow between ancestral populations, before the radiation of northern juncos took place. This suggests that *bairdi* received gene flow from the ancestral BPCO population during a long period. The analyses also recovered a high migration rate to a from the ancestral CO population to *bairdi*, and a reverse, one order of magnitude lower rate from *bairdi* to CO, a band somehow unlikely for such divergent lineages and their distant geographic locations. Similarly, *bairdi* presented signs of low but significant incoming gene flow from *phaeonotus*, *caniceps* and *oreganus*. These patterns may be an artifact resulting from the previous high migration rate from PCO to *bairdi*, which keeps appearing in derived populations due to extensive introgressions and high levels of shared variability. Nevertheless, signs of low gene flow among highly differentiated forms of *Junco* have been previously reported based on mitochondrial markers (Aleixandre et al. 2013), so G-PhoCS inferences of contemporary gene flow among divergent lineages should not be ruled out.

High values of divergence scattered across the genome based on both relative (F_{ST}) and absolute (D_{XY}) indexes supported the role of multiple selective factors acting in a high number of independent loci. Parallel estimates of F_{ST} and D_{XY} are useful for recovering a more confident signal of divergence mediated by selection (evidenced by positive correlation between F_{ST} and D_{XY}), or to detect high F_{ST} values merely due to low levels of within-group variability (evidenced by negative correlation between F_{ST} and D_{XY}) because of background selection or selective sweeps acting on linked loci (Cruickshank and Hahn 2014; but see Burri et al. 2015). In contrast, what we found was widespread highly differentiated loci and no obvious regions of correlated values between the two indices for both old-young and young-young comparisons. This also supports limited gene flow among current lineages, since substantial gene flow would yield a pattern of reduced D_{XY} across genome (Patterson et al. 2012).

Conclusions

The results of this final Chapter confirm the recent origin of the boreal forms of *Junco* and reinforce the hypothesis of a diversification process during a demographic expansion and postglacial recolonization of the North American continent. A method for demographic inference based on a full coalescent isolation-with-migration model applied to whole-genome data revealed a demographic history including moderate bottlenecks and subsequent population expansions, but limited gene flow among recent lineages, which suggests that the high phenotypic diversity of the dark-eyed junco originated in isolated populations that became established during the recolonization. Genomic surveys of Tajima's D and differentiation indices F_{ST} and D_{XY} based on GBS data also supported the demographic expansion with limited gene flow hypothesis, and revealed a genomic landscape congruent with a process of diversification driven by multiple selective factors acting over a high number of independent loci. These patterns corroborate most of the previous reported results, and support the role of combined effects of historical and selective factors promoting phenotypic divergence and lineage diversification in one of the fastest cases of speciation known in vertebrates.

References

Aleixandre, P., J. Hernández-Montoya, and B. Milá. 2013. Speciation on oceanic islands: rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: *Junco*). PLOS One 8(5):e63242.

Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, and J. Thibault. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current protocols in bioinformatics:11.10. 11-11.10. 33.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics:btu170.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, and L. Z. Garamszegi. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome research 25:1656-1665.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, and I. J. Lovette. 2017. Repeated divergent selection on pigmentation genes in a rapid finch radiation. Science Advances 3:e1602404.

Clark, A. G., M. J. Hubisz, C. D. Bustamante, S. H. Williamson, and R. Nielsen. 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome research 15:1496-1502.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler, and A. M. McCabe. 2009. The last glacial maximum. science 325:710-714.

Cruickshank, T. E. and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133-3157.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, and M. Hanna. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Anglais 43:491-498.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological Speciation: an integrative approach. Molecular ecology 23:513-521.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28:342-350.

Freedman, A. H., I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han, P. M. Silva, M. Galaverni, Z. Fan, P. Marx, and B. Lorente-Galdos. 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS genetics 10:e1004016.

Friis, G., P. Aleixandre, R. Rodríguez-Estrella, A. G. Navarro-Sigüenza, and B. Milá. 2016. Rapid postglacial diversification and long-term stasis within the songbird genus Junco: phylogeographic and phylogenomic evidence. Molecular Ecology 25:6175-6195.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Anglais 43:1031-U1151.

Jensen, J. D., A. Wong, and C. F. Aquadro. 2007. Approaches for identifying targets of positive selection. TRENDS in Genetics 23:568-577.

Krueger, F. 2015. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Laurent, S., S. P. Pfeifer, M. L. Settles, S. S. Hunter, K. M. Hardwick, L. Ormond, V. C. Sousa, J. D. Jensen, and E. B. Rosenblum. 2016. The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. Molecular ecology 25:306-323.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

Li, J., H. Li, M. Jakobsson, S. Li, P. SjÖDin, and M. Lascoux. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology 21:28-44.

Librado, P. and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

Mayr, E. 1963. Animal species and evolution. Belknap Press, Cambridge, MA.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20:1297-1303.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M. and W. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences U.S.A. 76:5269-5273.

Nolan, V. J., E. D. Ketterson, D. A. Cristol, C. M. Rogers, E. D. Clotfelter, R. C. Titus, S. J. Schoech, and E. Snajdr. 2002. Dark-eyed Junco (*Junco hyemalis*) *in* A. Poole, and F. Gill, eds. The Birds of North America. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Nosil, P. 2008. Speciation with gene flow could be common. Molecular Ecology 17:2103-2106.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:"isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. Ancient admixture in human history. Genetics 192:1065-1093.

R_Core_Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24:4348-4370.

Riesch, R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, T. E. Farkas, K. Lucek, E. Hellen, V. Soria-Carrasco, and S. R. Dennis. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. Nature Ecology & Evolution 1:0082.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Simpson, G. 1953. The major features of evolution. and G. Simpson, Tempo and Mode in:143.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585-595.

Turner, S. D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv:005165.

Via, S. 2009. Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106:9939-9946.

Warnes, G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, A. Liaw, T. Lumley, M. Maechler, A. Magnusson, and S. Moeller. 2013. gplots: Various R programming tools for plotting data. R package version 2.12. 1. Available online at: h ttp://CRAN. R-project. org/package= gplots.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358-1370.

GENERAL DISCUSSION AND CONCLUSIONS

Junco h. oreganus, picture by Roy Hancliff, www.royhancliff.com

GENERAL DISCUSSION

Recently diversified systems are optimal research subjects to study the mechanisms by which drift, gene flow and selection shape diversity and promote lineage formation (Ritchie 2007; Lerner et al. 2011). In the early stages of the speciation process, causal correlations between patterns of phenotypic divergence and genetic adaptive variability with specific selective factors are still recent and detectable (e.g. Losos et al. 1998; Kocher 2004; Rundell and Price 2009). In turn, reconstructing the evolutionary history of closely related lineages is a challenging task because of lack of resolution due to incomplete lineage sorting and potential gene flow (Cariou et al. 2013). In this thesis, we have implemented an experimental design that combines phylogeographic and phylogenomic data with phenotypic and ecological analyses, with the fundamental aim of studying mechanisms of early lineage divergence in the non-model biological system of the genus *Junco* in order to and draw robust, generalizable conclusions about the process of speciation.

The evolutionary history of the genus Junco*: a case of rapid diversification during the postglacial recolonization of North America*

The high diversity of *Junco* forms across the broad geographic distribution of the genus has raised numerous questions and debates both among evolutionary biologists and taxonomists about their evolutionary relationships and about how to classify them (Mayr 1942; Ketterson and Atwell 2016; Milá et al. 2016). A major aim of this thesis project was to reconstruct the evolutionary history of the genus and test the potential postglacial origin of the phenotypically divergent junco forms of North America proposed by Milá et al. (2007a). The results reported in the Chapter I revealed the existence of at least four genetically divergent lineages in Central America that have been geographically isolated in oceanic or sky islands for over two hundred thousand years; and a set of closely related lineages in USA and Canada, corresponding to the phenotypically differentiated, geographically structured northern forms of junco. Molecular analyses based on mtDNA confirmed the postglacial origin of the young northern lineages ca. 18,000 years ago, while phylogeographic analyses recovered clear signs of population expansion of dark-eyed junco forms across North America as evidenced by star-like haplotype networks and significant tests. In addition, phylogenomic reconstructions based on

genome-wide genotyping-by-sequencing (GBS, Elshire et al. 2011) markers resolved the polytomic relationships of specific and subspecific forms of yellow-eyed and dark-eyed juncos into reciprocally monophyletic groups, reinforcing the hypothesis of an extremely fast process of lineage diversification and phenotypic differentiation during the recolonization of the North American continent by one of the yellow-eyed lineages from southern Mexico. In sharp contrast, old Central American lineages presented relatively low levels of phenotypic differentiation, consistent with an anagenetic process of long-term stasis in geographic isolation. Overall, phenotypic diversity is seemingly correlated with the number of cladogenetic events across the *Junco* phylogenetic reconstruction and decoupled from lineage age, a pattern that has been shown to emerge at macroevolutionary scales (Rabosky and Adams 2012; Rabosky et al. 2012; Rabosky et al. 2013).

The case of the junco radiation exemplifies the importance of using genome-wide data when analyzing recent speciation processes. In North America, phylogeographic analyses based on mtDNA markers led to the early conclusion that most recent speciation events occurred through population divergence in isolated refugia during Pleistocene glacial maxima (Johnson and Cicero 2004; Weir and Schluter 2004; Lovette 2005). The specific process of demographic expansion and recolonization after glacial maximums as a potential driver of speciation has received much less attention (Milá et al. 2007a), and patterns of widespread sharing of single haplotypes among populations with certain degree of phenotypic or ecological differentiation in temperate regions have been commonly interpreted as evidence of panmixia and long-distance gene flow (e.g. Zink 1996; Oomen et al. 2011). However, the increased phylogenetic resolution provided by genome-wide SNPs has demonstrated that the lack of structure in mtDNA is due not to gene flow, but to very recent origin and shared ancestral polymorphism (Milá et al. 2006; Ruegg et al. 2006). Furthermore, the phylogenomic reconstruction recovered with remarkable precision the south-to-north sequence of cladogenetic events, from yellow-eyed forms in Mexico to the most recently evolved forms in the north, confirming the role of a rapid postglacial expansion in driving rapid diversification among populations with limited gene flow. These outcomes underscore the importance of combining mtDNA and nuclear markers in phylogeographic and

phylogenetic inferences (Edwards and Bensch 2009). High-throughput sequencing is demonstrating its strong potential to recover genome-wide informative markers to overcome problems of lack of resolution (Rubin et al. 2012; Cariou et al. 2013; Wagner et al. 2013), and in so doing previously signals of low structure among populations deemed the result of pervasive gene flow or mtDNA introgression are often shown to represent recent lineages with high levels of shared ancestral polymorphisms (Milá et al. 2007b; Brelsford and Irwin 2009; Milá et al. 2011; Ruegg et al. 2014).

Unlike other avian models in which recent divergence appears to be limited to a reduced number of regions or "genomic islands of divergence" (Ellegren et al. 2012; Poelstra et al. 2014), differentiation among junco forms based on SNP neutral markers indicates that divergence has taken place at the level of the entire genome. This suggests a role for local adaptation to environmental conditions along the latitudinal gradient, which is thought to result from relatively weak multifarious selection acting on many loci across the genome as opposed to strong selection on a few loci (Dambroski and Feder 2007; Michel et al. 2010; Parchman et al. 2013; Egea-Serrano et al. 2015), a hypothesis that we tested in Chapters III and IV (see below).

Further support for the hypothesis of recent diversification during a northward demographic expansion with limited gene flow among diverging populations came from the whole-genome analysis in Chapter IV. We used the recently developed program generalized phylogenetic coalescent sampler (G-PhoCS, Gronau et al. 2011) and seven fully sequenced genomes representing one of the old southern lineages (*bairdi*) and three of the young northern lineages (*phaeonotus*, *caniceps*, and *oreganus*). G-PhoCS estimates of divergence times were consistent with the inferences based on mtDNA markers and molecular clocks tested in BEAST. The analysis also revealed consistent patterns of recent demographic bottlenecks followed by periods of demographic expansion for the young lineages, and a sustained signal of strong demographic reduction for the isolated *bairdi* form. Measures of Tajima's D based on GBS genome-wide markers yielded congruent patterns of demographic evolution for all tested lineages. In addition, the G-PhoCS analysis recovered signs of ancient gene flow between *bairdi* and the ancestral

common ancestor of *phaeonotus*, *caniceps* and *oreganus*, but limited rates of post-divergence migration between the young lineages of *Junco*. These outcomes reinforce once again the importance of genome-wide data when analyzing the evolutionary and demographic history in recent speciation scenarios, and reflect considerable consistency in the estimated rate of divergence for genome-wide SNPs and mtDNA markers.

A role for sexual selection driving rapid phenotypic differentiation

The high rate of phenotypic change and marked differentiation in plumage pattern and color among recently radiated northern forms of junco, suggested that sexual selection might have played a relevant role in their diversification, a hypothesis that we tested in Chapter II. Mantel tests based on colorimetric data showed a significant correlation between the intensity of sexual selection —estimated by the degree of sexual dichromatism (Owens and Hartley 1998; Dunn et al. 2001; Huang and Rabosky 2014; Cooney et al. 2017), and the amount of divergence in plumage coloration, especially when controlling for neutral genetic differentiation. In contrast, discriminant function analyses on morphometric variables revealed limited divergence among northern juncos, suggesting relatively weaker selective pressures on ecomorphological traits than on traits potentially under sexual selection (Panhuis et al. 2001).

Rapid diversification of sexually selected traits across closely related species and populations has been documented in a number of taxa (e.g. Arnegard et al. 2010; Safran et al. 2013; Martin and Mendelson 2014), which suggests that sexual selection may act independently of ecological factors over isolated populations and be a more rapid promoter of phenotypic diversification at early stages of speciation (Price 1998; Kraaijeveld et al. 2011). However, and despite the relatively low divergence in ecomorphological traits, at least three aspects in the northern junco radiation suggest potential interactions between ecological factors and mate choice based on secondary sexual traits, as proposed by recent models suggesting the combined action of natural and sexual selection in diversification (van Doorn et al. 2009; Kraaijeveld et al. 2011; Maan and Seehausen 2011; Butlin et al. 2012; Wagner et al. 2012; Scordato et al. 2014). First, the diversification process in northern juncos

occurred during a demographic expansion along a latitudinal gradient, and major ecological differences are present across their distribution, potentially exerting considerable divergent selection. Second, as noted above, patterns of genome-wide divergence suggest the action of multiple selective factors driving differentiation on many loci across the genome, which contradicts a process of diversification limited to a few sexually selected loci such as those controlling plumage coloration. And third, our multivariate and linear regression analyses yielded statistical support to the latitudinal pattern of increasing sexual dichromatism from the divergent Central American lineages to the North American lineages, suggesting more intense sexual selection in the recently radiated boreal forms. Association of sexual dichromatism with breeding latitude and with migratory behavior is a well-documented pattern in birds (Badyaev and Hill 2003), and may reflect coupled processes of local adaptation and development of mate preference (Hamilton 1961; Fitzpatrick 1994; Badyaev and Hill 2003; Friedman et al. 2009).

Nevertheless, the interactions between divergent mate choice and ecological factors, and the relative contributions of sexual and natural selection to lineage divergence remain poorly understood and difficult to assess (Safran et al. 2013; Scordato et al. 2014). In addition, in the case of the juncos the latitudinal distribution of the postglacial boreal forms reflects not only the ecological gradient across which their demographic expansion occurred, but also the historical sequence of cladogenetic events that resulted in the multiple phylogenetic lineages of the complex. This implies that historical and demographic effects need to be taken into account when testing for selection driving diversification (Rellstab et al. 2015; Forester et al. 2016). In Chapter III of the present dissertation, we used redundancy analysis to test separately for associations between genetic variance and both sexual dichromatism and breeding latitude, but the test proved impossible to tease apart the effects of sexual selection in determining adaptive genetic variability from the effects of breeding latitude. This suggests that the loci potentially controlling the expression of sexual dimorphism may be highly structured in terms of latitude (Lasky et al. 2012). Conversely, it could mean that loci involved in adaptations to latitude-related factors are linked to the expression of sexual dichromatism. In any case, this outcome shows that breeding latitude (and ecologically associated

behaviors like seasonal migration) is a highly relevant parameter in determining the appearance of trait-based mate choice in juncos, and support a joint role for sexual and ecological factors in driving lineage differentiation.

Local adaptation along a selective gradient may be a factor in the development of mate choice behavior, but it does not explain how divergent mate choice appears under similar selective pressures (Scordato et al. 2014; Winger and Bates 2015). The mutation-order theory (Schluter 2009) provides a plausible explanation, in which isolated populations may have developed ecologically advantageous sexual signaling independently. The stochastic order in which mutations appear may thereby have fostered the rapid diversification of northern junco lineages in terms of plumage color and patterning, even under similar sexual selective pressures (Schluter 2009; Nosil and Flaxman 2011; Mendelson et al. 2014; Winger and Bates 2015). Whether or not diversification of sexually selected traits has resulted in sufficient barriers to reproduction, the analyses reported here provide strong support for sexual selection as a driver of the rapid radiation of northern juncos, likely interacting with ecological factors related to a diversification process along a latitudinal selective gradient.

Natural selection and drift drive differentiation among young lineages of Oregon Junco

Phylogenetic and neutral genetic structure analyses along with the patterns of phenotypic variation described in the first two Chapters of this thesis depicted a scenario of rapid demographic expansion, divergence in isolation and both natural and sexual selective pressures potentially involved in the recent radiation of the boreal forms of *Junco*. The main objective of Chapter III was to disentangle the relative strengths and interactions of neutral and ecological selective factors promoting rapid lineage divergence in the Oregon junco. To do so we applied redundancy analyses, a powerful genetic-environment association (GEA) method that allow to test specific ecological hypothesis of genetic diversification while controlling by distorting effects as population structure and demography.

GEA and neutral genetic analyses revealed a striking array of divergence modes within the Oregon junco complex. Geographically isolated forms from north Baja

California *townsendi* and *pontilis* presented high neutral genetic differentiation, yet low genetic-environment association values, suggesting that differentiation is due to drift under conditions of small population size and reduced gene flow. In contrast, the *pinosus* form from California presented considerable genetic structure but also a distinctive signal of genetic-environment association, a pattern of positive correlation between neutral and adaptive genetic variability consistent with a process of isolation-by-adaptation (IBA, Nosil et al. 2008; Funk et al. 2011). We also found signs of divergence along a continuous selective gradient with no obvious geographic barriers to gene flow across the distribution of the *thurberi* Oregon junco, distributed from south to north California; and especially the northernmost forms, from northern *thurberi* populations in Tahoe to the *oreganus* form in southern Alaska, showing great intergradation both spatially and in terms of neutral genetic differentiation yet considerable differentiation in the genetic-environment correlation patterns along a gradient of increasing humidity and vegetation cover.

The latitudinal range of the Oregon junco spans most of western North America, and the lineage has itself differentiated into distinct populations in a way that parallels the entire junco radiation, suggesting that similar processes may have been involved in the radiation of the complex. Immediate future prospects in junco speciation research include a thorough study of ecological factors of diversification in a historical framework to test to what extent these results can be extrapolated to the entire diversification of the system.

Genomic landscapes of differentiation support a process of genome-wide divergence driven by multifarious selection on multiple independent loci in the Dark-eyed junco radiation

In Chapters III and IV, we used GBS markers to study how signs of differentiation among juncos are structured across the genome and to infer patterns of selection-driven divergence. Because F_{ST} has been criticized as a reliable measure of divergent selection (Cruickshank and Hahn 2014), we resort to different genetic indexes and methods to correct for population history and intrinsic levels of genetic variation. In Chapter III, we used BayeScan to conduct the genomic surveys of Oregon junco genotypes mapped against the zebra finch. However, the zebra finch and the juncos

diverged around 20 million years ago, which results in many mismatches between the junco sequences and the reference genome. Thus, Oregon junco surveys are relatively less dense, and potentially biased towards more conserved regions of the genome. Even so, the genomic landscapes of variation among Oregon juncos revealed the presence of outliers putatively under positive selection scattered across the complete genome, seemingly supporting the hypothesis proposed in Chapter I and specifically tested in Chapter III of a potential role for multiple selective pressures driving genome-wide divergence due to local adaptation along the latitudinally broad and heterogeneous distribution of the dark-eyed juncos.

In Chapter IV, we used parallel measures of F_{ST} and D_{XY} to study genomic landscapes of divergence between yellow-eyed and dark-eyed juncos, and between the subspecific forms of dark-eyed junco *caniceps* and *oreganus*. Once again, we found no obvious regions of high differentiation, but outliers scattered all along the analyzed scaffolds for both F_{ST} and D_{XY} indices. Besides supporting the hypothesis of multifarious selection acting over genome-wide widespread loci also among more differentiated lineages of northern juncos, these outcomes are consisted with similar patterns reported in other taxa at early stages of the speciation process (e.g. Parchman et al. 2013; Brawand et al. 2014; Egan et al. 2015).

CONCLUSIONS

- The reconstruction of the evolutionary history of the genus *Junco* revealed a set of rapidly diversified young lineages in North America contrasting with at least four isolated lineages in the south showing relative phenotypic stasis. This pattern suggests that the rate of lineage formation is correlated with the rate of phenotypic change and not clade age.
- Phylogenetic and phylogenomic analyses confirmed the postglacial origin of northern junco forms and recovered a pattern of reciprocal monophyly among them, congruent with a process of northward recolonization of the North American continent from Mexico.
- Phylogenetic and population structure signal supported the genetic identity of phenotypically differentiated forms, congruent with a scenario of limited gene flow among independent lineages differentiated in allopatry during the recolonization.
- Whole-genome analyses in G-PhoCS also supported the recent origin of the boreal forms of *Junco*, and recovered signals of demographic expansions and limited gene flow during the diversification of young lineages.

- The northern juncos presented an extremely fast rate of phenotypic diversification in terms of plumage coloration that contrasted with minor differences in ecomorphological traits and correlated with the strength of sexual selection, estimated by the degree of sexual dichromatism.
- The genus *Junco* showed a pattern of increasing sexual dichromatism along the latitudinal ecological gradient, suggesting concomitant effects of natural and sexual selection. This hypothesis is consistent with the observed association of latitude of breeding range and sexual dichromatism with genome-wide adaptive genetic variance.
- Redundancy analyses revealed overlapping effects of both latitude of breeding range and sexual selection in shaping genome-wide adaptive variance, suggesting a role for divergent mate choice and ecological factors jointly driving lineage differentiation.

- The phenotypic radiation of the northern forms of *Junco* may have taken place under similar sexual selective pressures acting over isolated populations fostered by high evolvability of feather color patterns and the stochastic order in which mutations appeared.

- Neutral structure and genotype-environment association (GEA) analyses revealed that both neutral and selective factors have driven divergence among subspecific lineages of the Oregon junco. The combined effects of a demographic expansion along a selective gradient with a heterogeneous environmental landscape have resulted in a striking array of speciation modes including forms that have differentiated due to drift in isolated 'sky islands', ecologically isolated forms despite a lack of obvious geographic barriers to gene flow, and adaptive diversifications of intergraded forms along ecological clines.
- GEA and niche divergence tests supported a process of diversification driven by multiple ecological factors including elevation, precipitation and temperature during the breeding season, and different vegetation indexes.
- Genome surveys based on GBS single nucleotide polymorphisms in Oregon juncos were consistent with a diversification process driven by multiple ecological factors acting on many loci across the genome, what suggests that selection may promote local adaptation in short periods when driven by several factors acting over multiple independent loci.
- The signal of ecologically based differentiation found in the Oregon junco, whose latitudinal range is similar to that of all northern junco forms, suggests that local adaptation may have played a role in the diversification of the complete dark-eyed junco system.

- The North American juncos represent one of the fastest radiations known in vertebrates, and was driven by the combined effects of historical processes such as demographic expansions and drift in isolation, and also selective factors, including natural and sexual selection. The complex also represents a compelling example of a speciation continuum, harboring geographical forms that range from the locally differentiated population to

the species and from complete isolation to extensive hybridization. The complex array of patterns of variation and of speciation modes found in the genus *Junco* provides a unique opportunity to study speciation and this small but diverse group of taxa may well become a model system in the field of Evolutionary Biology.

CONCLUSIONES

- La reconstrucción de la historia evolutiva del género *Junco* reveló un conjunto de linajes jóvenes diversificados en un corto periodo de tiempo en Norte América, que contrastaban con los al menos cuatro linajes aislados encontrados en el sur de la distribución, que presentaban signos de estasis fenotípica en términos relativos. Este patrón sugiere que en los juncos la tasa de formación de nuevos linajes está correlacionado con la tasa de cambio fenotípico, y no con la edad del clado.
- Los análisis filogenéticos y filogenómicos confirmaron el origen postglacial de los morfotipos septentrionales, y recuperaron un patrón de monofilia recíproca entre ellos congruente con un proceso de recolonización del continente norte americano desde México hacia el Canadá.
- Tanto la señal filogenética como de estructura poblacional neutra apoyaron la identidad genética de los morfotipos fenotípicamente diferenciados de junco, en congruencia con un escenario de bajo flujo genético entre los linajes, que se habrían diferenciado en alopatría durante la recolonización de Norte América.
- Análisis basados en genomas completos realizados en G-PhoCS respaldaron el origen reciente de los morfotipos boreales de *Junco*, y recuperaron señales de expansión demográfica y bajo flujo genético durante su diversificación.

- Los juncos del norte presentaron una tasa de diversificación fenotípica extremadamente rápida en cuanto a coloración del plumaje, que contrasta con diferencias menores en rasgos ecomorfológicos y correlaciona con la intensidad de la selección sexual, estimada por el grado de dimorfismo sexual.
- El género *Junco* presentó un patrón de incremento del dicromatismo sexual a lo largo del gradiente ecológico latitudinal, sugiriendo efectos concomitantes de la selección natural y sexual. Esta hipótesis es coherente con la asociación detectada de la latitud de cría y de dicromatismo sexual con la variabilidad genética adaptativa a nivel genómico.

- Los análisis de redundancia revelaron efectos solapantes de la latitud de cría y de la selección sexual en la determinación de la variabilidad genómica adaptativa, sugiriendo un papel de factores ecológicos y de diferencias en la elección de pareja promoviendo conjuntamente la diferenciación de los linajes.
- La radiación fenotípica de los morfotipos septentrionales de *Junco* puede haber ocurrido bajo circunstancias similares de selección sexual actuando sobre poblaciones aisladas, propiciado por un alto 'evolvability' del color del plumaje y el orden estocástico en que las mutaciones subyacentes aparecieron.

- La estructura genética y los análisis de asociación genotipo-ambiente (GEA) revelaron que factores tanto neutros como selectivos han estado implicados en la divergencia de los linajes subespecíficos del junco de Oregón. Los efectos combinados de una expansión demográfica a lo largo de un gradiente selectivo con un paisaje ambiental heterogéneo han resultado en una sorprendente colección de modos de especiación incluyendo morfotipos aislados que se han diferenciado por deriva génica en 'sky islands'; morfotipos aislados ecológicamente a pesar de la ausencia de barreras geográficas obvias; y diversificaciones adaptativas de morfotipos cuyas poblaciones se solapan a lo largo de clinas ecológicas.
- Tests GEA y de divergencia de nicho respaldaron un proceso de diversificación dirigido por múltiples factores ecológicos incluyendo elevación, precipitaciones y temperatura durante el periodo de cría, y diferentes índices de cobertura vegetal.
- Patrones de variabilidad genómica basados en GBS SNPs de juncos de Oregón resultaron congruentes con un proceso de diversificación dirigido por múltiples factores ecológicos actuando sobre muchos loci a lo largo del genoma, sugiriendo que la selección puede promover adaptación local in cortos periodos de tiempo cuando varias fuerzas evolutivas actúan sobre múltiples loci independientes.
- La señal de diferenciación ecológica encontrada en el junco de Oregón, cuya distribución latitudinal se asemeja a la del conjunto de los morfotipos

septentrionales de junco, sugiere que la adaptación local puede haber jugado un papel relevante en la diversificación del sistema completo.

- Los juncos de Norte América representan una de las radiaciones más rápidas conocidas en vertebrados, promovida por la combinación de los efectos de procesos históricos como expansiones demográficas y deriva génica, y también factores selectivos que incluyen la selección natural y sexual. El complejo *Junco* también representa un convincente ejemplo de 'speciation continuum' que comprende formas geográficas que van desde poblaciones localmente diferenciadas hasta especies, y del aislamiento completo a la hibridación extensiva. La colección de complejos patrones de variación y modos de especiación encontrados en el género *Junco* proporciona una oportunidad única para el estudio de la especiación, y este pequeño pero diverso grupo de taxones bien podría convertirse en un sistema modelo en el campo de la Biología Evolutiva.

References

Arnegard, M. E., P. B. McIntyre, L. J. Harmon, M. L. Zelditch, W. G. Crampton, J. K. Davis, J. P. Sullivan, S. Lavoué, and C. D. Hopkins. 2010. Sexual signal evolution outpaces ecological divergence during electric fish species radiation. The American Naturalist 176:335-356.

Badyaev, A. V. and G. E. Hill. 2003. Avian sexual dichromatism in relation to phylogeny and ecology. Annual Review of Ecology, Evolution, and Systematics 34:27-49.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, and E. Bezault. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375.

Brelsford, A. and D. E. Irwin. 2009. Incipient speciation despite little assortative mating: the yellow-rumped warbler hybrid zone. Evolution 63:3050-3060.

Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, C. R. Castillo, W. Diao, M. E. Maan, S. Paolucci, and F. J. Weissing. 2012. What do we need to know about speciation? Trends in Ecology & Evolution 27:27-39.

Cariou, M., L. Duret, and S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3:846-852.

Cooney, C. R., J. A. Tobias, J. T. Weir, C. A. Botero, and N. Seddon. 2017. Sexual selection, speciation and constraints on geographical range overlap in birds. Ecology Letters.

Cruickshank, T. E. and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133-3157.

Dambroski, H. R. and J. L. Feder. 2007. Host plant and latitude-related diapause variation in *Rhagoletis pomonella*: a test for multifaceted life history adaptation on different stages of diapause development. Journal of Evolutionary Biology 20:2101-2112.

Dunn, P. O., L. A. Whittingham, and T. E. Pitcher. 2001. Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. Evolution 55:161-175.

Edwards, S. V. and S. Bensch. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Molecular Ecology 18:2930-2933.

Egan, S. P., G. J. Ragland, L. Assour, T. H. Powell, G. R. Hood, S. Emrich, P. Nosil, and J. L. Feder. 2015. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. Ecology letters 18:817-825.

Egea-Serrano, A., S. Hangartner, A. Laurila, and K. Räsänen. 2015. Multifarious selection through environmental change: acidity and predator-mediated adaptive divergence in the moor frog (*Rana arvalis*). Proceedings of the Royal Society B 281:20133266.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756-760.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379.

Fitzpatrick, S. 1994. Colourful migratory birds: evidence for a mechanism other than parasite resistance for the maintenance of'good genes' sexual selection. Proceedings of the Royal Society of London B: Biological Sciences 257:155-160.

Forester, B. R., M. R. Jones, S. Joost, E. L. Landguth, and J. R. Lasky. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular ecology 25:104-120.

Friedman, N. R., C. M. Hofmann, B. Kondo, and K. E. Omland. 2009. Correlated evolution of migration and sexual dichromatism in the New World orioles (Icterus). Evolution 63:3269-3274.

Funk, D. J., S. P. Egan, and P. Nosil. 2011. Isolation by adaptation in Neochlamisus leaf beetles: host-related selection promotes neutral genomic divergence. Molecular Ecology 20:4671-4682.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Anglais 43:1031-U1151.

Hamilton, T. H. 1961. On the functions and causes of sexual dimorphism in breeding plumage characters of North American species of warblers and orioles. The American Naturalist 95:121-123.

Huang, H. and D. Rabosky. 2014. Sexual Selection and Diversification: Reexamining the Correlation between Dichromatism and Speciation Rate in Birds. American Naturalist, The 184:E101-E114.

Johnson, N. K. and C. Cicero. 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution 58:1122-1130.

Ketterson, E. D. and J. W. Atwell. 2016. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco. University of Chicago Press.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nature reviews. Genetics 5:288.

Kraaijeveld, K., F. J. L. Kraaijeveld-Smit, and M. E. Maan. 2011. Sexual selection and speciation: the comparative evidence revisited. Biological Reviews 86:367-377.

Lasky, J. R., D. L. Des Marais, J. McKAY, J. H. Richards, T. E. Juenger, and T. H. Keitt. 2012. Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Molecular Ecology 21:5512-5529.

Lerner, H. R. L., M. Meyer, H. F. James, M. Hofreiter, and R. C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers. Current Biology 21:1838-1844.

Losos, J. B., T. R. Jackman, A. Larson, K. de Quiroz, and L. Rodriguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115-2118.

Lovette, I. J. 2005. Glacial cycles and the tempo of avian speciation. Trends in Ecology and Evolution 20:57-59.

Maan, M. E. and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecology letters 14:591-602.

Martin, M. D. and T. C. Mendelson. 2014. Changes in sexual signals are greater than changes in ecological traits in a dichromatic group of fishes. Evolution 68:3618-3628.

Mayr, E. 1942. Speciation in the Junco. Ecology 23:378-379.

Mendelson, T. C., M. D. Martin, and S. M. Flaxman. 2014. Mutation-order divergence by sexual selection: diversification of sexual signals in similar environments as a first step in speciation. Ecology letters 17:1053-1066.

Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread genomic divergence during sympatric speciation. Proceedings of the National Academy of Sciences 107:9724-9729.

Milá, B., P. Aleixandre, S. Alvarez-Nordström, J. McCormack, E. Ketterson, and J. Atwell. 2016. More than meets the eye: Lineage diversity and evolutionary history of Dark-eyed and Yellow-eyed juncos. Snowbird: Integrative Biology and Evolutionary Diversity in the Junco (ED Ketterson and JW Atwell, Editors). University of Chicago Press, Chicago, Illinois, USA:179-198.

Milá, B., J. E. McCormack, G. Castaneda, R. K. Wayne, and T. B. Smith. 2007a. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus Junco. Proceedings of the Royal Society B-Biological Sciences 274:2653-2660.

Milá, B., T. B. Smith, and R. K. Wayne. 2006. Postglacial population expansion drives the evolution of long-distance migration in a songbird. Evolution 60:2403-2409.

Milá, B., T. B. Smith, and R. K. Wayne. 2007b. Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. Molecular Ecology 16:159-173.

Milá, B., D. P. L. Toews, T. B. Smith, R. K. Wayne, J. Lake, and N. F. Dixie. 2011. A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (Aves : *Dendroica coronata*). Biological Journal of the Linnean Society 103:696-706.

Miller, A. 1941. Speciation in the avian genus *Junco*. University of California Publications in Zoology 44:173-434.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes:"isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316-336.

Nosil, P. and S. M. Flaxman. 2011. Conditions for mutation-order speciation. Proceedings of the Royal Society of London B: Biological Sciences 278:399-407.

Oomen, R. A., M. W. Reudink, J. J. Nocera, C. M. Somers, M. C. Green, and C. J. Kyle. 2011. Mitochondrial evidence for panmixia despite perceived barriers to gene flow in a widely distributed waterbird. Journal of Heredity 102:584-592.

Owens, I. P. and I. R. Hartley. 1998. Sexual dimorphism in birds: why are there so many different forms of dimorphism? Proceedings of the Royal Society of London B: Biological Sciences 265:397-407.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends in Ecology and Evolution 16:325-413.

Parchman, T., Z. Gompert, M. Braun, R. Brumfield, D. McDonald, J. Uy, G. Zhang, E. Jarvis, B. Schlinger, and C. Buerkle. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology 22:3304-3317.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410-1414.

Price, T. 1998. Sexual selection and natural selection in bird speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 353:251-260.

Rabosky, D., F. Santini, J. Eastman, S. Smith, B. Sidlauskas, J. Chang, and M. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature Communications 4:1958.

Rabosky, D. L. and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. Evolution 66:1807-1818.

Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade Age and Species Richness Are Decoupled Across the Eukaryotic Tree of Life. PLoS Biol 10:e1001381.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology 24:4348-4370.

Ritchie, M. G. 2007. Sexual selection and speciation. Pp. 79-102. Annual Review of Ecology, Evolution, and Systematics.

Rubin, B. E. R., R. H. Ree, and C. S. Moreau. 2012. Inferring Phylogenies from RAD Sequence Data. Plos One 7:e33394-e33394.

Ruegg, K., E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Molecular Ecology 23:4757-4769.

Ruegg, K. C., R. J. Hijmans, and C. Moritz. 2006. Climate change and the origin of migratory pathways in the Swainson's thrush, *Catharus ustulatus*. Journal of Biogeography 33:1172-1182.

Rundell, R. J. and T. D. Price. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution 24:394-399.

Safran, R. J., E. S. Scordato, L. B. Symes, R. L. Rodríguez, and T. C. Mendelson. 2013. Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. Trends in ecology & evolution 28:643-650.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323:737-741.

Scordato, E. S., L. B. Symes, T. C. Mendelson, and R. J. Safran. 2014. The role of ecology in speciation by sexual selection: a systematic empirical review. Journal of Heredity 105:782-794.

van Doorn, G. S., P. Edelaar, and F. J. Weissing. 2009. On the origin of species by natural and sexual selection. Science 326:1704-1707.

Wagner, C., L. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366-369.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22:787-798.

Weir, J. T. and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. Proceedings of the Royal Society Biological Sciences Series B 271:1881-1887.

Winger, B. M. and J. M. Bates. 2015. The tempo of trait divergence in geographic isolation: avian speciation across the Marañon Valley of Peru. Evolution 69:772-787.

Zink, R. M. 1996. Comparative phylogeography in North American birds. Evolution 50:308-317.

ANNEX I: R scripts

R code for computing the avian visual model variables for chapters I and II based on colorimetric measures from museum specimens

```
### Color treatment ###
setwd('D:/PhD/5. Papers/3. DEJU_sexual/2017_02_22 Sexual_reset/4.
RDA/8. RDA_WTSP8_bayescan')

library(pavo)

##-------------------------------------------------------------------------
# Extraction of spectra and AV variables
##-------------------------------------------------------------------------

jspecs <- getspec( 'D:/PhD/5. Papers/3. DEJU_sexual/2017_02_22
Sexual_reset/1. Muestrario',
                   ext="txt", decimal=".", subdir = T, subdir.names =
T, lim = c(300, 700), fast = T)


#explorespec(jspecs[,c(1:15)], by=3, lwd=2)

jsplit1 <- strsplit(names(jspecs), "/")
vecagg <- array(NA, length(names(jspecs)))

jsplit2 = jsplit1
for (i in 2:length(names(jspecs))){
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00000" |
jsplit2[[i]][3] == "Spec00001" | jsplit2[[i]][3] == "Spec00002",
"Crown", jsplit2[[i]][3])
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00003" |
jsplit2[[i]][3] == "Spec00004" | jsplit2[[i]][3] == "Spec00005",
"Nape", jsplit2[[i]][3])
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00006" |
jsplit2[[i]][3] == "Spec00007" | jsplit2[[i]][3] == "Spec00008",
"Back", jsplit2[[i]][3])
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00009" |
jsplit2[[i]][3] == "Spec00010" | jsplit2[[i]][3] == "Spec00011",
"Breast", jsplit2[[i]][3])
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00012" |
jsplit2[[i]][3] == "Spec00013" | jsplit2[[i]][3] == "Spec00014",
"Belly", jsplit2[[i]][3])
  jsplit2[[i]][3] <- ifelse(jsplit2[[i]][3] == "Spec00015", "Flank",
jsplit2[[i]][3])
}

for (i in 2:length(names(jspecs))){
  vecagg[i] <- paste(jsplit2[[i]][1], jsplit2[[i]][2],
jsplit2[[i]][3], sep="_")
}

jmspecs <- aggspec(jspecs, by=vecagg[-1], FUN=mean)

#---- For VULCANI
jspecs.vulc <- getspec( 'D:/PhD/5. Papers/3. DEJU_sexual/2017_02_22
Sexual_reset/1.1 VULCANI',
```

```
                    ext="txt", decimal=".", subdir = T, subdir.names =
T, lim = c(300, 700), fast = T)

jsplit1.vulc <- strsplit(names(jspecs.vulc), "/")
vecagg.vulc <- array(NA, length(names(jspecs.vulc)))

jsplit2.vulc = jsplit1.vulc
for (i in 2:length(names(jspecs.vulc))){
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00000" |
jsplit2.vulc[[i]][3] == "Spec00001" | jsplit2.vulc[[i]][3] ==
"Spec00002", "Crown", jsplit2.vulc[[i]][3])
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00003" |
jsplit2.vulc[[i]][3] == "Spec00004" | jsplit2.vulc[[i]][3] ==
"Spec00005", "Nape", jsplit2.vulc[[i]][3])
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00006" |
jsplit2.vulc[[i]][3] == "Spec00007" | jsplit2.vulc[[i]][3] ==
"Spec00008"  | jsplit2.vulc[[i]][3] == "Spec00009" |
jsplit2.vulc[[i]][3] == "Spec00010" | jsplit2.vulc[[i]][3] ==
"Spec00011", "Back", jsplit2.vulc[[i]][3])
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00012" |
jsplit2.vulc[[i]][3] == "Spec00013" | jsplit2.vulc[[i]][3] ==
"Spec00014", "Breast", jsplit2.vulc[[i]][3])
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00015" |
jsplit2.vulc[[i]][3] == "Spec00016" | jsplit2.vulc[[i]][3] ==
"Spec00017", "Belly", jsplit2.vulc[[i]][3])
  jsplit2.vulc[[i]][3] <- ifelse(jsplit2.vulc[[i]][3] == "Spec00018",
"Flank", jsplit2.vulc[[i]][3])
}

for (i in 2:length(names(jspecs.vulc))){
  vecagg.vulc[i] <- paste(jsplit2.vulc[[i]][1], jsplit2.vulc[[i]][2],
jsplit2.vulc[[i]][3], sep="_")
}

jmspecs.vulc <- aggspec(jspecs.vulc, by=vecagg.vulc[-1], FUN=mean)


jmspecs.all <- merge(jmspecs, jmspecs.vulc)
jmspecs.all.fix <- procspec(jmspecs.all, opt='smooth', span = 0.1,
fixneg='addmin')

##------------------------------------------------------------------------
# Reordering
##------------------------------------------------------------------------
jmspecs.table <- t(jmspecs.all.fix)
colnames(jmspecs.table) <- jmspecs.table[1,]
jmspecs.table <- jmspecs.table[-1,]

sample.data <-
as.data.frame(t(as.data.frame(strsplit(row.names(jmspecs.table),
'_'))))
colnames(sample.data) <- c('Morph', 'ID', 'Patch')
row.names(sample.data) <- c()

sample.data$Sex <-
as.data.frame(sapply(strsplit(as.character(sample.data$ID), ""), tail,
1))
colnames(sample.data[,4]) <- 'Sex'
row.names(sample.data) <- paste(sample.data$Morph, sample.data$ID,
sample.data$Patch, sep = '_')
```

```
spectral.data <- merge(sample.data, jmspecs.table, by = 'row.names')
spectral.data <- spectral.data[,-1]

spectral.data$ID <- gsub('[A-Z]', '', spectral.data$ID)
spectral.data$ID <- factor(as.character(spectral.data$ID))

nlevels(spectral.data$ID)

#---- wl_patch matrix

spectral.dataset <- matrix(nrow = 665, ncol = 3+401*6 )

dimnames(spectral.dataset) <- list(NULL,c("ID","Morph","Sex",
                                  paste(300:700,
levels(spectral.data$Patch)[1], sep="_"),
                                  paste(300:700,
levels(spectral.data$Patch)[2], sep="_"),
                                  paste(300:700,
levels(spectral.data$Patch)[3], sep="_"),
                                  paste(300:700,
levels(spectral.data$Patch)[4], sep="_"),
                                  paste(300:700,
levels(spectral.data$Patch)[5], sep="_"),
                                  paste(300:700,
levels(spectral.data$Patch)[6], sep="_")))

spectral.dataset[,1] <- as.character(levels(factor(spectral.data$ID)))

for (i in 1:nrow(spectral.data)) {

  line <- match(as.character(spectral.data[i,2]),
as.character(spectral.dataset[,1]))
  variables <- paste(300:700, spectral.data[i,3], sep="_")
  spectral.dataset[line,variables] <-
as.character(spectral.data[i,5:405])

}

match <- match(spectral.dataset[,1],spectral.data[,2])
spectral.dataset[,2] <- as.character(spectral.data$Morph[match])
spectral.dataset[,3] <- as.character(spectral.data$Sex[match])

write.table(spectral.dataset, 'CompleteSpectra_bypatch.txt', row.names
= F, col.names = T, sep = '\t', quote = F)

##-----------------------------------------------------------------------
# Computing AVM variables
##-----------------------------------------------------------------------
cone.sim <- vismodel(jmspecs.all.fix,
                     visual = "bluetit",
                     achromatic = "bt.dc",
                     illum = "forestshade",
                     trans = "bluetit",
                     qcatch = "Qi",
                     bkg = "ideal",
                     vonkries = T,
                     relative = T)

visual.avm <- colspace(cone.sim, space = "auto")
norm.bright <- as.data.frame(colSums(jmspecs.all.fix[,2:3991])/401)
```

```
colnames(norm.bright) <- 'Norm_Bright'
visual <- merge(visual.avm, norm.bright, by = 'row.names')
row.names(visual) <- visual$Row.names
visual <- visual[,-1]

visual.data <-
as.data.frame(t(as.data.frame(strsplit(row.names(visual), '_'))))
colnames(visual.data) <- c('Morph', 'ID', 'Patch')
row.names(visual.data) <- c()

visual.data$Sex <-
as.data.frame(sapply(strsplit(as.character(visual.data$ID), ""), tail,
1))
colnames(visual.data[,4]) <- 'Sex'
row.names(visual.data) <- paste(visual.data$Morph, visual.data$ID,
visual.data$Patch, sep = '_')


avm.variables <- merge(visual.data, visual, by = 'row.names')
avm.variables <- avm.variables[,-1]

avm.variables$ID <- gsub('[A-Z]', '', avm.variables$ID)
avm.variables$ID <- factor(as.character(avm.variables$ID))

nlevels(avm.variables$ID)

#---- Ordering dataset
avm.dataset <- matrix(nrow = 665, ncol = 3+17*6 )

dimnames(avm.dataset) <- list(NULL,c("ID","Morph","Sex",
                                     paste(colnames(visual),
levels(avm.variables$Patch)[1], sep="_"),
                                     paste(colnames(visual),
levels(avm.variables$Patch)[2], sep="_"),
                                     paste(colnames(visual),
levels(avm.variables$Patch)[3], sep="_"),
                                     paste(colnames(visual),
levels(avm.variables$Patch)[4], sep="_"),
                                     paste(colnames(visual),
levels(avm.variables$Patch)[5], sep="_"),
                                     paste(colnames(visual),
levels(avm.variables$Patch)[6], sep="_")))

avm.dataset[,1] <- as.character(levels(factor(avm.variables$ID)))


for (i in 1:nrow(avm.variables)) {

  line <- match(as.character(avm.variables[i,2]),
as.character(avm.dataset[,1]))
  variables <- paste(colnames(visual), avm.variables[i,3], sep="_")
  avm.dataset[line,variables] <- as.character(avm.variables[i,5:21])

}

match <- match(avm.dataset[,1],avm.variables[,2])
avm.dataset[,2] <- as.character(avm.variables$Morph[match])
avm.dataset[,3] <- as.character(avm.variables$Sex[match])

write.table(avm.dataset, 'CompleteAVM_bypatch.txt', row.names = F,
col.names = T, sep = '\t', quote = F)
```

```
#---- Only informative AVM variables
avm.4varsdataset <- avm.dataset[,1:3]

for (i in 0:5) {

  avm.4varsdataset <- cbind(avm.4varsdataset, avm.dataset[
,c((3+(12+(i*17))), (3+(13+(i*17))), (3+(16+(i*17))),
(3+(17+(i*17)))))])

}

write.table(avm.4varsdataset, 'AVM_4variables_CompleteSet.txt',
row.names = F, col.names = T, quote = F, sep = '\t')
```

R code for linear regression and discriminant function analyses for sex and for forms of Junco *based on colorimetric data; the principal components analysis based on genome-wide SNPs; and the redundancy analyses based on genome-wide SNPs and colorimetric data applied in Chapter II.*

```
### Color treatment ###
setwd('D:/PhD/5. Papers/3. DEJU_sexual/4. RDA/4.
FourthRound_STRUdataset')

library(MASS)
library(xlsx)
library(SNPRelate)
library(vegan)
library(ggfortify)
library(klaR) # Detach after use, masks rda vegan function


       #****************** COLOR ANALYSES ******************#


##---------------------------------------------------------------------
#Buiding Color Dataset and removing per group outliers
##---------------------------------------------------------------------
avm4.df <- read.table('AVM_4variables_COMPLETE.txt', header = T)
avm4.df[,1:3] <- lapply(avm4.df[,1:3], as.character)

avm4.split.list <- split(avm4.df, avm4.df$Species)

## Remove outliers with boxplot
for (i in 1:length(avm4.split.list)) {
  for (k in 4:27) {

    outliers <- boxplot.stats(avm4.split.list[[i]][[k]], coef =
10)$out
    avm4.split.list[[i]][[k]] <- replace(avm4.split.list[[i]][[k]],
avm4.split.list[[i]][[k]] %in% outliers, NA)

  }
}


avm4.rmout.df <- do.call('rbind', avm4.split.list)
avm4.rmout.df <- na.omit(avm4.rmout.df)
row.names(avm4.rmout.df) <- seq(1:nrow(avm4.rmout.df))
avm4.rmout.df$Species <- as.factor(avm4.rmout.df$Species)

##---------------------------------------------------------------------
# Stepwise Sex DFA ALL SPECIES
##---------------------------------------------------------------------

## Stepwise variable selection and DFA
##---------------------------------------------------------------------
sexdfa.all.stepwise <- greedy.wilks(Sex ~ h.theta_Back + h.phi_Back +
r.achieved_Back + Norm_Bright_Back
                                    + h.theta_Belly + h.phi_Belly +
r.achieved_Belly + Norm_Bright_Belly
```

```
                                + h.theta_Breast + h.phi_Breast +
r.achieved_Breast + Norm_Bright_Breast
                                + h.theta_Crown + h.phi_Crown +
r.achieved_Crown + Norm_Bright_Crown
                                + h.theta_Flank + h.phi_Flank +
r.achieved_Flank + Norm_Bright_Flank
                                + h.theta_Nape + h.phi_Nape +
r.achieved_Nape + Norm_Bright_Nape
                                , avm4.rmout.df)


step.allsexdfa.variables <-
as.vector(sexdfa.all.stepwise$results[[1]])
stepvar.length <- as.numeric(length(step.allsexdfa.variables))
step.allsexdfa.variables[(stepvar.length+1):(stepvar.length+3)] <-
c('ID', 'Species', 'Sex')
step.allsexdfa.variables <-
step.allsexdfa.variables[c((stepvar.length+1):(stepvar.length+3),
1:(stepvar.length))]


avm4.allsexdfastep.df <- avm4.rmout.df[,step.allsexdfa.variables]
avm4.allsexdfastep.df$Species <- factor(avm4.allsexdfastep.df$Species)
str(avm4.allsexdfastep.df)

step.allsexdfa <- lda(Sex ~.,
avm4.allsexdfastep.df[,3:ncol(avm4.allsexdfastep.df)])

step.allsexdfa.LD <- as.data.frame(predict(step.allsexdfa)$x[,1])
colnames(step.allsexdfa.LD) <- 'DF_1'
avm4.allsexdfastep_DF1.df <- cbind(avm4.allsexdfastep.df,
step.allsexdfa.LD)

avm4.allsexdfastep_DF1.df$DF_meanCR <- vector(mode = 'numeric', length
= nrow(avm4.allsexdfastep_DF1.df ))
avm4.allsexdfastep_DF1.list <- split(avm4.allsexdfastep_DF1.df ,
avm4.allsexdfastep_DF1.df$Species)


for (i in 1:length(unique(avm4.allsexdfastep_DF1.df$Species))) {

  meancorrect.list <-
by(avm4.allsexdfastep_DF1.list[[i]][[ncol(avm4.allsexdfastep_DF1.df)-
1]],
                         avm4.allsexdfastep_DF1.list[[i]][[3]], mean)

  meancorrect.factor <- ((as.numeric(meancorrect.list[2]) -
as.numeric(meancorrect.list[1]))/2) + as.numeric(meancorrect.list[1])

  avm4.allsexdfastep_DF1.list[[i]][[ncol(avm4.allsexdfastep_DF1.df)]]
<-
as.numeric(avm4.allsexdfastep_DF1.list[[i]][[ncol(avm4.allsexdfastep_D
F1.df)-1]]) - meancorrect.factor

}

avm4.allsexdfastep_DF1CR.df <- do.call('rbind',
avm4.allsexdfastep_DF1.list)
avm4.allsexdfastep_DF1CR.df <-
avm4.allsexdfastep_DF1CR.df[order(avm4.allsexdfastep_DF1CR.df$Species)
,]
```

```
row.names(avm4.allsexdfastep_DF1CR.df) <-
seq(1:nrow(avm4.allsexdfastep_DF1CR.df))

summary(avm4.allsexdfastep_DF1.df$Species)
write.table(avm4.allsexdfastep_DF1CR.df[,c(2:3,
ncol(avm4.allsexdfastep_DF1CR.df))], 'DisfactorCR_allsexdfastep.txt',
            row.names = F, col.names = T, sep = '\t', quote = F)

## Computing Dimorphism scores
##---------------------------------------------------------------------

# Males
colormales.allsexstepdfa.df <-
avm4.allsexdfastep_DF1CR.df[avm4.allsexdfastep_DF1CR.df$Sex == 'M',
c(2, 4:ncol(avm4.allsexdfastep_DF1CR.df))]
colormales.allsexstepdfa.mean <-
aggregate(colormales.allsexstepdfa.df[,
2:ncol(colormales.allsexstepdfa.df)],

list(colormales.allsexstepdfa.df$Species), mean)

colnames(colormales.allsexstepdfa.mean)[1] <- 'Species'
row.names(colormales.allsexstepdfa.mean) <-
colormales.allsexstepdfa.mean$Species
colormales.allsexstepdfa.mean <-
colormales.allsexstepdfa.mean[order(row.names(colormales.allsexstepdfa
.mean)), ]

# Females
colorfemales.allsexstepdfa.df <-
avm4.allsexdfastep_DF1CR.df[avm4.allsexdfastep_DF1CR.df$Sex == 'F',
c(2, 4:ncol(avm4.allsexdfastep_DF1CR.df))]
colorfemales.allsexstepdfa.mean <-
aggregate(colorfemales.allsexstepdfa.df[,
2:ncol(colorfemales.allsexstepdfa.df)],

list(colorfemales.allsexstepdfa.df$Species), mean)

colnames(colorfemales.allsexstepdfa.mean)[1] <- 'Species'
row.names(colorfemales.allsexstepdfa.mean) <-
colorfemales.allsexstepdfa.mean$Species
colorfemales.allsexstepdfa.mean <-
colorfemales.allsexstepdfa.mean[order(row.names(colorfemales.allsexste
pdfa.mean)), ]

# Dichromatism score and Dim dataset computation
dimscore.allsexstepdfa.df <-
as.data.frame(abs(colorfemales.allsexstepdfa.mean$DF_meanCR -
colormales.allsexstepdfa.mean$DF_meanCR))
names(dimscore.allsexstepdfa.df) <- 'Dim_score'
row.names(dimscore.allsexstepdfa.df) <-
row.names(colorfemales.allsexstepdfa.mean)

dimset.allsexstepdfa.df <-
as.data.frame(abs(colorfemales.allsexstepdfa.mean[,2:ncol(colorfemales
.allsexstepdfa.mean)] -
colormales.allsexstepdfa.mean[,2:ncol(colormales.allsexstepdfa.mean)])
)
rownames(dimset.allsexstepdfa.df) <-
row.names(dimscore.allsexstepdfa.df)
```

```
## Linear regression with LAT
##------------------------------------------------------------------------
gbs.alllat.centroids <- read.table('GBSsamples_latcentroids.txt',
header = T)
gbs.alllat.centroids <-
gbs.alllat.centroids[gbs.alllat.centroids$Species %in%
row.names(dimscore.allsexstepdfa.df),]

dimscore.allsexstepdfa.lat <- cbind(dimscore.allsexstepdfa.df,
gbs.alllat.centroids)
str(dimscore.allsexstepdfa.lat)

plot(dimscore.allsexstepdfa.lat$Dim_score ~
dimscore.allsexstepdfa.lat$LAT)
abline(lm(Dim_score ~ LAT, dimscore.allsexstepdfa.lat))
text(dimscore.allsexstepdfa.lat$LAT,
     dimscore.allsexstepdfa.lat$Dim_score+0.05,
     labels = dimscore.allsexstepdfa.lat$Species)

dimlat.allsexstepdfa.reg <- lm(Dim_score ~ LAT,
dimscore.allsexstepdfa.lat)
summary(dimlat.allsexstepdfa.reg)
cor(dimscore.allsexstepdfa.lat$Dim_score,
dimscore.allsexstepdfa.lat$LAT,
    method = 'pearson')

##------------------------------------------------------------------------
# Stepwise Sex DFA RADIATION
------------------------------------------------------------------

## Stepwise variable selection and DFA
##------------------------------------------------------------------------
sexdfa.rad.stepwise <- greedy.wilks(Sex ~ h.theta_Back + h.phi_Back +
r.achieved_Back + Norm_Bright_Back
                                    + h.theta_Belly + h.phi_Belly +
r.achieved_Belly + Norm_Bright_Belly
                                    + h.theta_Breast + h.phi_Breast +
r.achieved_Breast + Norm_Bright_Breast
                                    + h.theta_Crown + h.phi_Crown +
r.achieved_Crown + Norm_Bright_Crown
                                    + h.theta_Flank + h.phi_Flank +
r.achieved_Flank + Norm_Bright_Flank
                                    + h.theta_Nape + h.phi_Nape +
r.achieved_Nape + Norm_Bright_Nape
                                    , avm4.radrmout.df)


step.radsexdfa.variables <-
as.vector(sexdfa.rad.stepwise$results[[1]])
stepvar.length <- as.numeric(length(step.radsexdfa.variables))
step.radsexdfa.variables[(stepvar.length+1):(stepvar.length+3)] <-
c('ID', 'Species', 'Sex')
step.radsexdfa.variables <-
step.radsexdfa.variables[c((stepvar.length+1):(stepvar.length+3),
1:(stepvar.length))]


avm4.radsexdfastep.df <- avm4.radrmout.df[,step.radsexdfa.variables]
avm4.radsexdfastep.df$Species <- factor(avm4.radsexdfastep.df$Species)
str(avm4.radsexdfastep.df)
```

```
step.radsexdfa <- lda(Sex ~.,
avm4.radsexdfastep.df[,3:ncol(avm4.radsexdfastep.df)])

step.radsexdfa.LD <- as.data.frame(predict(step.radsexdfa)$x[,1])
colnames(step.radsexdfa.LD) <- 'DF_1'
avm4.radsexdfastep_DF1.df <- cbind(avm4.radsexdfastep.df,
step.radsexdfa.LD)

avm4.radsexdfastep_DF1.df$DF_meanCR <- vector(mode = 'numeric', length
= nrow(avm4.radsexdfastep_DF1.df ))
avm4.radsexdfastep_DF1.list <- split(avm4.radsexdfastep_DF1.df ,
avm4.radsexdfastep_DF1.df$Species)


for (i in 1:length(unique(avm4.radsexdfastep_DF1.df$Species))) {

  meancorrect.list <-
by(avm4.radsexdfastep_DF1.list[[i]][[ncol(avm4.radsexdfastep_DF1.df)-
1]],
                          avm4.radsexdfastep_DF1.list[[i]][[3]], mean)

  meancorrect.factor <- ((as.numeric(meancorrect.list[2]) -
as.numeric(meancorrect.list[1]))/2) + as.numeric(meancorrect.list[1])

  avm4.radsexdfastep_DF1.list[[i]][[ncol(avm4.radsexdfastep_DF1.df)]]
<-
as.numeric(avm4.radsexdfastep_DF1.list[[i]][[ncol(avm4.radsexdfastep_D
F1.df)-1]]) - meancorrect.factor

}

avm4.radsexdfastep_DF1CR.df <- do.call('rbind',
avm4.radsexdfastep_DF1.list)
avm4.radsexdfastep_DF1CR.df <-
avm4.radsexdfastep_DF1CR.df[order(avm4.radsexdfastep_DF1CR.df$Species)
,]
row.names(avm4.radsexdfastep_DF1CR.df) <-
seq(1:nrow(avm4.radsexdfastep_DF1CR.df))

summary(avm4.radsexdfastep_DF1.df$Species)
write.table(avm4.radsexdfastep_DF1CR.df[,c(2:3,
ncol(avm4.radsexdfastep_DF1CR.df))], 'DisfactorCR_radsexdfastep.txt',
            row.names = F, col.names = T, sep = '\t', quote = F)

## Computing Dimorphism scores
##-----------------------------------------------------------------------

# Males
colormales.radsexstepdfa.df <-
avm4.radsexdfastep_DF1CR.df[avm4.radsexdfastep_DF1CR.df$Sex == 'M',
c(2, 4:ncol(avm4.radsexdfastep_DF1CR.df))]
colormales.radsexstepdfa.mean <-
aggregate(colormales.radsexstepdfa.df[,
2:ncol(colormales.radsexstepdfa.df)],

list(colormales.radsexstepdfa.df$Species), mean)

colnames(colormales.radsexstepdfa.mean)[1] <- 'Species'
row.names(colormales.radsexstepdfa.mean) <-
colormales.radsexstepdfa.mean$Species
```

```
colormales.radsexstepdfa.mean <-
colormales.radsexstepdfa.mean[order(row.names(colormales.radsexstepdfa
.mean)), ]

# Females
colorfemales.radsexstepdfa.df <-
avm4.radsexdfastep_DF1CR.df[avm4.radsexdfastep_DF1CR.df$Sex == 'F',
c(2, 4:ncol(avm4.radsexdfastep_DF1CR.df))]
colorfemales.radsexstepdfa.mean <-
aggregate(colorfemales.radsexstepdfa.df[,
2:ncol(colorfemales.radsexstepdfa.df)],

list(colorfemales.radsexstepdfa.df$Species), mean)

colnames(colorfemales.radsexstepdfa.mean)[1] <- 'Species'
row.names(colorfemales.radsexstepdfa.mean) <-
colorfemales.radsexstepdfa.mean$Species
colorfemales.radsexstepdfa.mean <-
colorfemales.radsexstepdfa.mean[order(row.names(colorfemales.radsexste
pdfa.mean)), ]

# Dichromatism score and Dim dataset computation
dimscore.radsexstepdfa.df <-
as.data.frame(abs(colorfemales.radsexstepdfa.mean$DF_meanCR -
colormales.radsexstepdfa.mean$DF_meanCR))
names(dimscore.radsexstepdfa.df) <- 'Dim_score'
row.names(dimscore.radsexstepdfa.df) <-
row.names(colorfemales.radsexstepdfa.mean)

dimset.radsexstepdfa.df <-
as.data.frame(abs(colorfemales.radsexstepdfa.mean[,2:ncol(colorfemales
.radsexstepdfa.mean)] -
colormales.radsexstepdfa.mean[,2:ncol(colormales.radsexstepdfa.mean)])
)
rownames(dimset.radsexstepdfa.df) <-
row.names(dimscore.radsexstepdfa.df)

gbs.radlat.centroids <- read.table('GBSsamples_latcentroids.txt',
header = T)
gbs.radlat.centroids <-
gbs.radlat.centroids[gbs.radlat.centroids$Species %in%
row.names(dimscore.radsexstepdfa.df),]
dimscore.radsexstepdfa.lat <- cbind(dimscore.radsexstepdfa.df,
gbs.radlat.centroids)


##------------------------------------------------------------------
# Stepwise Species DFA RADIATION
##------------------------------------------------------------------
avm4.radrmout_males.df <- avm4.radrmout.df[avm4.radrmout.df$Sex=='M',]
speciesdfa.rad.stepwise <- greedy.wilks(Species ~ h.theta_Back +
h.phi_Back + r.achieved_Back + Norm_Bright_Back
                                        + h.theta_Belly + h.phi_Belly
+ r.achieved_Belly + Norm_Bright_Belly
                                        + h.theta_Breast +
h.phi_Breast + r.achieved_Breast + Norm_Bright_Breast
                                        + h.theta_Crown + h.phi_Crown
+ r.achieved_Crown + Norm_Bright_Crown
                                        + h.theta_Flank + h.phi_Flank
+ r.achieved_Flank + Norm_Bright_Flank
```

```
                                          + h.theta_Nape + h.phi_Nape +
r.achieved_Nape + Norm_Bright_Nape
                                          , avm4.radrmout_males.df)


step.radspeciesdfa.variables <-
as.vector(speciesdfa.rad.stepwise$results[[1]])
stepvar.length <- as.numeric(length(step.radspeciesdfa.variables))
step.radspeciesdfa.variables[(stepvar.length+1):(stepvar.length+3)] <-
c('ID', 'Species', 'Sex')
step.radspeciesdfa.variables <-
step.radspeciesdfa.variables[c((stepvar.length+1):(stepvar.length+3),
1:(stepvar.length))]


avm4.radspeciesdfastep.df <-
avm4.radrmout_males.df[,step.radspeciesdfa.variables]
avm4.radspeciesdfastep.df$Species <-
factor(avm4.radspeciesdfastep.df$Species)
str(avm4.radspeciesdfastep.df)

step.radspeciesdfa <- lda(Species ~., avm4.radspeciesdfastep.df[,c(2,
4:ncol(avm4.radspeciesdfastep.df))])

col.rainbow <- rainbow(nlevels(avm4.radrmout_males.df$Species))
palette(col.rainbow)

avm4.radspeciesdfastep.LD1<-predict(step.radspeciesdfa)$x[,1]
avm4.radspeciesdfastep.LD2<-predict(step.radspeciesdfa)$x[,2]
radspeciesdfastep.centroids <-
aggregate(cbind(avm4.radspeciesdfastep.LD1,
avm4.radspeciesdfastep.LD2) ~ Species, avm4.radrmout_males.df, mean)
colnames(radspeciesdfastep.centroids) <- c('Species', 'LD1', 'LD2')

plot(-avm4.radspeciesdfastep.LD2, avm4.radspeciesdfastep.LD1, col =
col.rainbow[avm4.radrmout_males.df$Species], pch = 16)
points(-radspeciesdfastep.centroids$LD2,
radspeciesdfastep.centroids$LD1, col = 'black', pch = 15)
legend("bottomleft", legend=levels(avm4.radrmout_males.df$Species),
pch= 16, col=1:nlevels(avm4.radrmout_males.df$Species))



          #****************** SNPS ANALYSES ******************#


source('D:/PhD/3. Dats&Anls/Scripts/GenomeScans/plot_R.r',
encoding='UTF-8')
detach(package: klaR)


##-----------------------------------------------------------------------
# SNP treatment: RADIATION BayeScan and PCA
##-----------------------------------------------------------------------



# BayeScan SNP sets construction
##-----------------------------------------------------------------------
```

```
mydata <-
read.table('GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B2
94.sel', colClasses = 'numeric')
snps.fst <-
read.table('GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B2
94_fst.txt',
                       colClasses = 'numeric')

# Plot setting the False Discovery Rate threshold to 0.1

results <-
plot_bayescan('GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001
_B294_fst.txt',
                          FDR=0.3)

length(results$outliers)
snps.fst.outliers <- snps.fst[results$outliers,]
snps.fst.outliers$SNP_ID <- row.names(snps.fst.outliers)

snps.fst.selected <- snps.fst.outliers[snps.fst.outliers$alpha>0,]
snps.fst.balanced <- snps.fst.outliers[snps.fst.outliers$alpha<0,]

# Write vcf file, remember to restitute first lines of vcf output file
and create the tab delimited one with TASSEL

vcf.table <-
read.table("GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B2
94.recode.vcf", header = F)
vcf.table.neutral <- vcf.table[-(results$outliers), ]

write.table(vcf.table.neutral,
            file =
'GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B294_neutral0
3.vcf',
            quote = F, sep = '\t', row.names = F, col.names = F)

vcf.table.selected <- vcf.table[snps.fst.selected$SNP_ID, ]
write.table(vcf.table.selected,
            file =
'GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B294_selected
03.vcf',
            quote = F, sep = '\t', row.names = F, col.names = F)


## LD pruning and PCA
#-----------------------------------------------------------------------
vcf.fn <-
'GBSradMAX8 biall dp2100 q70 maf003 maxmiss0901 hwe00001 B294 neutral0
3.vcf'
snpgdsVCF2GDS (vcf.fn, "snpsradNEU03.gds", method="copy.num.of.ref")
snpgdsSummary("snpsradNEU03.gds")

genofile <- openfn.gds('snpsradNEU03.gds')
snpset <- snpgdsLDpruning(genofile, ld.threshold=0.2, autosome.only =
FALSE)

snpset.id <- unlist(snpset)
sample.id <- read.gdsn(index.gdsn(genofile, "sample.id"))

# Write neutral, unlinked SNPs for other analyses
```

```
write.table (snpset.id, file = 'LEsnps_ID.txt', quote = F, sep = '\t',
row.names = F, col.names = F)

vcf.table <-
read.table("GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B2
94_neutral03.vcf",
                          header = F)

snps.id <- read.table('LEsnps_ID.txt')

vcf.LEsnps <- vcf.table[snps.id$V1, ]
write.table (vcf.LEsnps, file =
'GBSradMAX8_biall_dp2100_q70_maf003_maxmiss0901_hwe00001_B294_neutral0
3_LEsnps02.vcf',
             quote = F, sep = '\t', row.names = F, col.names = F)

# Load a file with the pops and performing pca
samples.dataset <- subset(gbs.dataset.comp, GBS.name %in% sample.id)

samples.sorted <- samples.dataset[order(samples.dataset$GBS.name), ]
samples.sorted.unique <- subset (samples.sorted,
!duplicated(samples.sorted$GBS.name))

pop_code <- samples.sorted.unique$Species
pop_code <- factor(pop_code)
unique(pop_code)

pca <- snpgdsPCA(genofile, sample.id = NULL, snp.id = snpset.id,
autosome.only = FALSE,
                 remove.monosnp = TRUE, maf = NaN, missing.rate = NaN,
eigen.cnt = 32,
                 num.thread = 1, bayesian = FALSE, need.genmat =
FALSE,
                 genmat.only = FALSE, verbose = TRUE)

# Saving eigenvectors for plotting
tab <- data.frame(sample.id = pca$sample.id,
                  pop = factor(pop_code)[match(pca$sample.id,
sample.id)],
                  EV1 = pca$eigenvect[,1],
                  EV2 = pca$eigenvect[,2],
                  EV3 = pca$eigenvect[,3],
                  EV4 = pca$eigenvect[,4],
                  EV5 = pca$eigenvect[,5],
                  EV6 = pca$eigenvect[,6],
                  EV7 = pca$eigenvect[,7],
                  EV8 = pca$eigenvect[,8],
                  EV9 = pca$eigenvect[,9],
                  EV10 = pca$eigenvect[,10],
                  stringsAsFactors = FALSE)

# Writting out EVs for other analyses
write.table (tab, file = 'Eigenvectors.txt', row.names = F, sep =
'\t', quote = F)

tab <- read.table('Eigenvectors.txt', header = T)
# Setting the palette of colors
col.rainbow <- rainbow(nlevels(tab$pop))
palette(col.rainbow)
```

```
# Plotting 2 PCs
plot(x=-tab$EV2, y=-tab$EV1, col= as.integer(tab$pop), xlab="PC2",
ylab="PC1", pch = 16)
legend("bottomleft", legend=levels(tab$pop), col=1:nlevels(tab$pop),
pch = 16)

# PCA_centroids
rad.centroids.pca <- aggregate(cbind(tab$EV1, tab$EV2, tab$EV3) ~ pop,
tab, mean)
row.names(rad.centroids.pca) <- rad.centroids.pca$pop
points(x=-rad.centroids.pca$V2, y=-rad.centroids.pca$V1, col= 'Black',
pch = 15)

# Plotting multiple PCs and explained percentage
pc.percent <- 100 * pca$eigenval[1:4]/sum(pca$eigenval, na.rm=T)
pc.percent

lbls <- paste("PC", 1:4, "\n", format(pc.percent[1:4], digits=2), "%",
sep="")
pairs(pca$eigenvect[,1:4], col=tab$pop, labels=lbls)



            #****************** RDA ******************#


# STEP VARIABLES RDA with Selected SNPs: dimscore and lat, variance
partition
##------------------------------------------------------------------------
rda_sel.radstepvariables <- rda(snps_SELrad.dataset.freq ~.,
dimscore.radsexstepdfa.df)

summary(rda_sel.radstepvariables)
coef(rda_sel.radstepvariables)
capture.output(summary(rda_sel.radstepvariables), file =
'Report_rda_Dim&LAT_dimscore_selsnps.txt')

rda_sel.radstepvariables
rda_sel.radstepvariables[[7]][[9]]

spe.part <- varpart(snps_SELrad.dataset.freq,
dimscore.radsexstepdfa.lat$Dim_score, dimscore.radsexstepdfa.lat$LAT)
plot(spe.part)
```

R code for the principal components analysis and the redundancy analyses from Chapter III based on genome-wide SNPs and ecological data

```
library(adegenet)
library(vegan)
library(SNPRelate)
library(gdsfmt)
library(xlsx)
library(seqRFLP)
library(packfor)

#-----------------------------------------------------------------------
# Bayescan filtering
#-----------------------------------------------------------------------

# Loading source code provided with Bayescan software
source('D:/PhD/3. Dats&Anls/Scripts/plot_R.r', encoding='UTF-8')

# Load Bayescan output files
mydata <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001.sel', colClasses = 'numeric')
snps.fst <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001_fst.txt',
                        colClasses = 'numeric')

# Plot setting the False Discovery Rate threshold to 0.3

results <-
plot_bayescan('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholeh
we00001_fst.txt',
                           FDR=0.3)

results$nb_outliers
head(results$outliers)

snps.fst.selected <- snps.fst[results$outliers,]
length(snps.fst.selected$alpha>0)

#------------- Write vcf file
# remember to restitute first lines of vcf output file

vcf.table <-
read.table("ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001.recode.vcf", header = F)
vcf.table <- vcf.table[-(results$outliers), ]

write.table(vcf.table,
            file =
'ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe00001_neutra
l03.vcf',
            quote = F, sep = '\t', row.names = F, col.names = F)

#-----------------------------------------------------------------------
## LD pruning and PCA
#-----------------------------------------------------------------------
```

```
vcf.fn <-
'ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe00001_neutra
l03.vcf'
snpgdsVCF2GDS (vcf.fn, "snpsNEU.gds", method="copy.num.of.ref")
snpgdsSummary("snpsNEU.gds")

genofile <- openfn.gds('snpsNEU.gds')
snpset <- snpgdsLDpruning(genofile, ld.threshold=0.2, autosome.only =
FALSE)

snpset.id <- unlist(snpset)
sample.id <- read.gdsn(index.gdsn(genofile, "sample.id"))

# Write neutral, unlinked SNPs for other analyses

write.table (snpset.id, file = 'LEsnps_ID.txt', quote = F, sep = '\t',
row.names = F, col.names = F)

vcf.table <-
read.table("ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001_neutral03.vcf",
                           header = F)
snps.id <- read.table('LEsnps_ID.txt')

vcf.LEsnps <- vcf.table[snps.id$V1, ]
write.table (vcf.LEsnps, file =
'ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe00001_neutra
l03_LEsnps02.vcf',
             quote = F, sep = '\t', row.names = F, col.names = F)


# Load a file with the pops and performing pca

gbs.dataset <- read.xlsx ('2016_03_03 GBS Junco
dataset_complete.xlsx', sheetName = 'Sheet1',
                          header = T, dec='.', as.data.frame=T)
samples.dataset <- subset(gbs.dataset, GBS.name %in% sample.id)

samples.sorted <- samples.dataset[order(samples.dataset$GBS.name), ]
samples.sorted.unique <- subset (samples.sorted,
!duplicated(samples.sorted$GBS.name))

pop_code <- samples.sorted.unique$Species

pca <- snpgdsPCA(genofile, sample.id = NULL, snp.id = snpset.id,
autosome.only = FALSE,
                 remove.monosnp = TRUE, maf = NaN, missing.rate = NaN,
eigen.cnt = 32,
                 num.thread = 1, bayesian = FALSE, need.genmat =
FALSE,
                 genmat.only = FALSE, verbose = TRUE)

# Saving eigenvectors for plotting
tab <- data.frame(sample.id = pca$sample.id,
                  pop = factor(pop_code)[match(pca$sample.id,
sample.id)],
                  EV1 = pca$eigenvect[,1],
                  EV2 = pca$eigenvect[,2],
                  EV3 = pca$eigenvect[,3],
                  EV4 = pca$eigenvect[,4],
                  EV5 = pca$eigenvect[,5],
```

```
                    EV6 = pca$eigenvect[,6],
                    EV7 = pca$eigenvect[,7],
                    EV8 = pca$eigenvect[,8],
                    EV9 = pca$eigenvect[,9],
                    EV10 = pca$eigenvect[,10],
                    stringsAsFactors = FALSE)

# Writting out EVs for other analyses
write.table (tab, file = 'Eigenvectors.txt', row.names = F, sep =
'\t', quote = F)

tab <- read.table('Eigenvectors.txt', header = T)
# Setting the palette of colors
col.rainbow <- rainbow(nlevels(tab$pop))
palette(col.rainbow)


# Plotting 2 PCs
color.figure <- read.table('colors_figure.txt', header = T, row.names
= 1, sep = '\t')
colvec <- rgb(red = color.figure, names = row.names(color.figure),
maxColorValue = 255)

plot(x=tab$EV2, y=-tab$EV1, xlab="PC2", ylab="PC1")
points(x=tab$EV2, y=-tab$EV1, col = colvec[tab$pop], pch = 21, bg =
colvec[tab$pop])
legend("bottomright", legend=levels(tab$pop), col = colvec, pch= 21,
pt.bg = colvec)


# Plotting multiple PCs and explained percentage
pc.percent <- 100 * pca$eigenval[1:4]/sum(pca$eigenval, na.rm=T)
pc.percent

lbls <- paste("PC", 1:4, "\n", format(pc.percent[1:4], digits=2), "%",
sep="")
pairs(pca$eigenvect[,1:4], col = colvec[tab$pop], labels=lbls, pch =
21, bg = colvec[tab$pop])


#------------------------------------------------------------------------
## Load 012 SNPs
#------------------------------------------------------------------------
# Whole Dataset

snps.dataset <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001.012',
                           header = F, row.names = 1, sep = '\t',
na.strings = '-1')
snps.indv <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001.012.indv')
row.names(snps.dataset) <- snps.indv$V1

snps.dataset <- snps.dataset[order(row.names(snps.dataset)), ]

snps.dataset.patt <- as.data.frame((matrix(0, nrow =
nrow(snps.dataset), ncol = ncol(snps.dataset))))

for (j in 1:ncol(snps.dataset)) {
```

```
  for (i in 1:nrow(snps.dataset)) {
#  print(paste('Patterson for SNP', j))
  u <- sum(snps.dataset[,j])/length(snps.dataset[,j])
  snps.dataset.patt[i,j] <- (snps.dataset[i,j] - u)/(((u/2)*(1-
(u/2)))^(1/2))
#  print(paste('Transformed SNP =', snps.dataset.patt[i,j]))
  }
  }

row.names(snps.dataset.patt) <- row.names(snps.dataset)

#-----------------------------------------------------------------------
## Load data
#-----------------------------------------------------------------------
geometrics <- read.table('2016_12_22
ORJUall_B294GBS_coords_clim_RAWfmiss.txt', header = T, row.names = 1,
sep = '\t')
geometrics <- geometrics[order(geometrics$Species),]
geometrics <- na.omit(geometrics[row.names(snps.dataset),])
geometrics <- geometrics[order(row.names(geometrics)), ]

# Keeping climatic variables:
clim <- geometrics[ ,c(9:32)]
clim <- clim[order(row.names(clim)), ]

#-----------------------------------------------------------------------
## RDA for the ecological hypothesis
#-----------------------------------------------------------------------
# Forward selection method
names(clim)
clim.ENM <- clim[,c(1, 12, 10, 18, 3, 4, 15, 24, 20, 21, 23)]
clim.ENM <- as.data.frame(scale(clim.ENM, center = T, scale = T))
names((clim.ENM))

rda.climENM <- rda(snps.dataset.patt ~., clim.ENM)
R2a.ENM <- RsquareAdj(rda.climENM)$adj.r.squared
forward.blanchet.ENM <- forward.sel(snps.dataset.patt, clim.ENM,
adjR2thresh = R2a.ENM, Xscale = F, Yscale = F)

forward.blanchet.ENM$variables
forward.blanchet.ENM

## Simple RDA
rda.climENM.stepf <- rda(snps.dataset.patt ~ tree_r + bio_10 + bio_18
+ srtm + ndvi_r3 + bio_3, clim)

vif.cca(rda.climENM.stepf)

anova.cca(rda.climENM.stepf, step = 10000)
RsquareAdj(rda.climENM.stepf)
summary(rda.climENM.stepf)

plot(rda.climENM.stepf)
points(rda.climENM.stepf, display = "sites", col =
colvec[geometrics$Species], pch = 21, bg = colvec[geometrics$Species])
legend("bottomright", legend = levels(geometrics$Species), col =
colvec, pch = 21, pt.bg = colvec)

climENMstepf.summary <- summary(rda.climENM.stepf)
capture.output(climENMstepf.summary, file
='climENMstepf_RDA_summaryreport.txt')
```

```
## Partial RDA
rda.climENMstepf.neutral <- rda(snps.dataset.patt ~ tree_r + bio_10 +
bio_18
                                + srtm + ndvi_r3 + bio_3
                                + Condition(tab$EV1 +tab$EV2),
data=clim)

anova.cca(rda.climENMstepf.neutral, step = 10000)
RsquareAdj(rda.climENMstepf.neutral)
summary(rda.climENMstepf.neutral)

plot(rda.climENMstepf.neutral)
points(rda.climENMstepf.neutral, display = "sites", col =
colvec[geometrics$Species], pch = 21, bg = colvec[geometrics$Species])
legend("bottomright", legend = levels(geometrics$Species), col =
colvec, pch = 21, pt.bg = colvec)

climENMstepfneutral.summary <- summary(rda.climENMstepf.neutral)
capture.output(climENMstepfneutral.summary, file
='climENMstepf_partialRDA_summaryreport.txt')


## Variance partition
clim.ENMstepf <- clim[, c(3, 10, 18, 20, 23, 24)]
clim.ENMstepf <- as.data.frame(scale(clim.ENMstepf, center = T, scale
= T))

neutral.vectors <- tab[ ,c(3, 4)]
row.names(neutral.vectors) <- tab$sample.id

spe.part <- varpart(snps.dataset.patt, clim.ENMstepf, neutral.vectors)
spe.part
plot(spe.part)

# Tests of all testable fractions
# Test of fractions [a+b] All variance explained by clim, equivalent
to the simple RDA
anova.cca(rda.climENM.stepf, step=10000)

# Test of fractions [b+c] All variance explained by coords
anova.cca(rda(snps.dataset.patt, neutral.vectors), step=10000)

# Test of fractions [a+b+c] All variance explained by both clim and
coords
env.neutralvecs <- cbind(clim.ENMstepf, neutral.vectors)
anova.cca(rda(snps.dataset.patt, env.neutralvecs), step=10000)

# Test of fraction [A] Variance explained only by clim, equivalent to
the partial RDA
anova.cca(rda.climENMstepf.neutral, step=10000)

# Test of fraction [C] Variance explained only by genetic structure
anova(rda(snps.dataset.patt, neutral.vectors, clim.ENMstepf),
step=10000)


#-----------------------------------------------------------------------
## RDA for the ecological hypothesis with BayeScan outliers
#-----------------------------------------------------------------------
vcf.table <-
```

```
read.table("ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001.recode.vcf", header = F)
vcf.selected <- vcf.table[(results$outliers), ]

write.table(vcf.selected,
            file =
'ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe00001_sel03.
vcf',
            quote = F, sep = '\t', row.names = F, col.names = F)

# Restore header and Convert to 012 with vcftools
# Load and transform 012

snps.divergent <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001_sel03.012',
                           header = F, row.names = 1, sep = '\t',
na.strings = '-1')
snps.indv <-
read.table('ORJU075x12geos_biall_dp2100_q40_nomissing_maf002_wholehwe0
0001_sel03.012.indv')
row.names(snps.divergent) <- snps.indv$V1

snps.divergent <- snps.divergent[order(row.names(snps.divergent)), ]

snps.divergent.patt <- as.data.frame((matrix(0, nrow =
nrow(snps.divergent), ncol = ncol(snps.divergent))))

for (j in 1:ncol(snps.divergent)) {
  for (i in 1:nrow(snps.divergent)) {
    #  print(paste('Patterson for SNP', j))
    u <- sum(snps.divergent[,j])/length(snps.divergent[,j])
    snps.divergent.patt[i,j] <- (snps.divergent[i,j] - u)/(((u/2)*(1-
(u/2)))^(1/2))
    #  print(paste('Transformed SNP =', snps.divergent.patt[i,j]))
  }
}

row.names(snps.divergent.patt) <- row.names(snps.divergent)

## Simple RDA
rda.climstepf.divergent <- rda(snps.divergent.patt~ srtm + ndvi_r3 +
bio_10 + bio_3 + bio_18 + tree_r, clim)

anova.cca(rda.climstepf.divergent, step = 10000)
RsquareAdj(rda.climstepf.divergent)
summary(rda.climstepf.divergent)

plot(rda.climstepf.divergent)
points(rda.climstepf.divergent, display = "sites", col =
colvec[geometrics$Species], pch = 21, bg = colvec[geometrics$Species])
legend("bottomright", legend = levels(geometrics$Species), col =
colvec, pch = 21, pt.bg = colvec)

climstepf.divergent.df <- summary(rda.climstepf.divergent)
capture.output(climstepf.divergent.df, file
='BayeScanOutliers_RDA_summaryreport.txt')
```

R code for vcf *files manipulation, Tajima's D boxplots and genome scans used in Chapter IV*

```
library(gplots)
library(plyr)
library(ggplot2)

#---------------------------------------------------------------------
# Build .vcfs for DnaSP
#---------------------------------------------------------------------
vcf.table <-
read.table("ORJU8_dp2100_q90_missing075_maf00625_hwe00001_hwe00001_B29
4.recode.vcf", header = F)

colnames(vcf.table)[1] <- 'CHROM'
colnames(vcf.table)[2] <- 'POS'
vcf.data <- vcf.table
vcf.table <- vcf.table[,1:2]

CHROM_ord <- vector(mode="numeric", length=0)
i <- 1
CHROM_ord[1] <- i

for (j in 2:length(vcf.table$CHROM)) {
  if (vcf.table$CHROM[j] == vcf.table$CHROM[j-1]) {
    CHROM_ord[j] <- i
  } else { i <- i+1
  CHROM_ord[j] <- i
  }
}

vcf.table <- cbind(vcf.table, CHROM_ord)
vcf.table <- vcf.table[order(vcf.table$CHROM_ord, vcf.table$POS), ]


## Create loci index
vcf.table$loci <- as.character(vcf.table$CHROM_ord)
chrom.groups <- split(vcf.table, vcf.table$CHROM_ord)

for (k in 1:length(chrom.groups)) {

  i <- 2
  j <- 1
  scaffold <- 1
  chrom.groups[[k]][[4]][[1]] <- paste(k, '_', scaffold, sep = '')

  for (h in 1:length(chrom.groups[[k]][[2]])) {

    if (i > length(chrom.groups[[k]][[2]])) {break

    } else {

      if (chrom.groups[[k]][[2]][[i]] - chrom.groups[[k]][[2]][[j]] >
68) {
        scaffold <- scaffold + 1
        chrom.groups[[k]][[4]][[i]] <- paste(k, '_', scaffold, sep =
'')
        i <- i+1
        j <- j+1
```

```
      } else {

        repeat {
          chrom.groups[[k]][[4]][[i]] <- paste(k, '_', scaffold, sep =
'')
          i <- i + 1
          if (i > length(chrom.groups[[k]][[2]]) ||
chrom.groups[[k]][[2]][[i]] - chrom.groups[[k]][[2]][[j]] > 68) {
            if (i > length(chrom.groups[[k]][[2]])) {break} else {

              scaffold <- scaffold +1
              chrom.groups[[k]][[4]][[i]] <- paste(k, '_', scaffold,
sep = '')
              j <- i
              i <- i+1
              break
            }}}}}}}


vcf.table.loci <- do.call('rbind', chrom.groups)
length(unique(vcf.table.loci$loci))
vcf.table$loci <- as.numeric(factor(vcf.table$loci))

## Redo vcf
vcf.complete <- merge(vcf.table.loci, vcf.data)
vcf.complete <- vcf.complete[order(vcf.complete$CHROM_ord,
vcf.complete$loci), ]

vcf.loci <- vcf.complete[,c(4, 2, 5:ncol(vcf.complete))]

write.table(vcf.loci,
            file =
'ORJU8_dp2100_q90_missing075_maf00625_hwe00001_hwe00001_B294_loci69.vc
f',
            quote = F, sep = '\t', row.names = F, col.names = F)


#----------------------------------------------------------------------
# Tajima's D
#----------------------------------------------------------------------
baju.tajima <- na.omit(read.table("BAJUtajima_D.txt", header = T,
na.strings = 'n.a.'))
yeju.tajima <- na.omit(read.table("YEJUtajima_D.txt", header = T,
na.strings = 'n.a.'))
ghju.tajima <- na.omit(read.table("GHJUtajima_D.txt", header = T,
na.strings = 'n.a.'))
orju.tajima <- na.omit(read.table("ORJUtajima_D.txt", header = T,
na.strings = 'n.a.'))

boxplot2(baju.tajima$TajimaD, yeju.tajima$TajimaD,
ghju.tajima$TajimaD, orju.tajima$TajimaD, top = T,
        notch = T, col = c("darkgoldenrod1","dodgerblue2"),
        main="Tajima's D Boxplots", names = c('bairdi', 'phaeonotus',
'caniceps', 'oreganus'))


#----------------------------------------------------------------------
# Dxy and Fst genome scans
#----------------------------------------------------------------------
library(qqman)
```

```
## Load vcf

vcf.table <-
read.table("CO_SNPs_dp2100_q70_missing075_hwe00001_maf004_bwa_WTSP.rec
ode.vcf", header = F)

colnames(vcf.table)[1] <- 'CHROM'
colnames(vcf.table)[2] <- 'POS'
vcf.data <- vcf.table
vcf.table <- vcf.table[,1:2]

## Load and reformat SNP table
WTSP.table <- read.table("Zonotrichia_albicollis.fa.fai", header = F)
WTSP.table <- WTSP.table[,1:2]

colnames(WTSP.table) <- c('CHROM', 'Length')
vcf.table <- merge(vcf.table, WTSP.table)
vcf.table <- vcf.table[order(vcf.table$Length, decreasing = T), ]

CHROM_ord <- vector(mode="numeric", length=0)
i <- 1
CHROM_ord[1] <- i

for (j in 2:length(vcf.table$CHROM)) {
  if (vcf.table$CHROM[j] == vcf.table$CHROM[j-1]) {
    CHROM_ord[j] <- i
  } else { i <- i+1
  CHROM_ord[j] <- i
  }
}

vcf.table <- cbind(vcf.table, CHROM_ord)
vcf.table <- vcf.table[order(vcf.table$CHROM_ord, vcf.table$POS), ]

## Create loci index
vcf.table$loci <- as.character(vcf.table$CHROM_ord)
vcf.table <- vcf.table[vcf.table$Length>2000000,]
length(unique(vcf.table$CHROM_ord))

chrom.groups <- split(vcf.table, vcf.table$CHROM_ord)



for (k in 1:length(chrom.groups)) {

  i <- 2
  j <- 1
  scaffold <- 1
  chrom.groups[[k]][[5]][[1]] <- paste(k, '_', scaffold, sep = '')

  for (h in 1:length(chrom.groups[[k]][[2]])) {

    if (i > length(chrom.groups[[k]][[2]])) {break

    } else {

      if (chrom.groups[[k]][[2]][[i]] - chrom.groups[[k]][[2]][[j]] >
68) {
        scaffold <- scaffold + 1
```

```
        chrom.groups[[k]][[5]][[i]] <- paste(k, '_', scaffold, sep =
'')
        i <- i+1
        j <- j+1

      } else {

        repeat {
          chrom.groups[[k]][[5]][[i]] <- paste(k, '_', scaffold, sep =
'')
          i <- i + 1
          if (i > length(chrom.groups[[k]][[2]]) ||
chrom.groups[[k]][[2]][[i]] - chrom.groups[[k]][[2]][[j]] > 68) {
            if (i > length(chrom.groups[[k]][[2]])) {break} else {

            scaffold <- scaffold +1
            chrom.groups[[k]][[5]][[i]] <- paste(k, '_', scaffold, sep
= '')
            j <- i
            i <- i+1
            break
    }}}}}}}


vcf.table.loci <- do.call('rbind', chrom.groups)
length(unique(vcf.table.loci$loci))

## Redo vcf
vcf.complete <- merge(vcf.table.loci, vcf.data)
vcf.complete <- vcf.complete[order(vcf.complete$CHROM_ord,
vcf.complete$POS), ]

vcf.loci <- vcf.complete[,c(5, 2, 6:ncol(vcf.complete))]

write.table(vcf.loci,
            file =
'CO_SNPs_dp2100_q70_missing075_hwe00001_maf004_bwa_WTSP_loci69.vcf',
            quote = F, sep = '\t', row.names = F, col.names = F)

## Loading Dxy after DnaSP and manhattan
Dxy.table <- na.omit(read.table('Dxy.txt', header = T, sep = '\t',
na.strings = 'n.a'))
loci.pos <- read.table('loci_pos.txt', header = T, sep = '\t')
Dxy.table <- merge(Dxy.table, loci.pos)
Dxy.table <- Dxy.table[order(Dxy.table$SNP),]
Dxy.table$SNP <- factor(Dxy.table$SNP)

col.topo <- topo.colors(2, alpha=1)
palette(col.topo)

## Outlier detection
boxplot(Dxy.table$Dxy)
dxy.outliers <- boxplot.stats(Dxy.table$Dxy, coef = 3)$out
dxy.outliers.index <- factor(Dxy.table[Dxy.table$Dxy %in%
dxy.outliers, ('SNP')])

fst.outliers <- boxplot.stats(Dxy.table$Fst, coef = 3)$out
fst.outliers.index <- factor(Dxy.table[Dxy.table$Fst %in%
fst.outliers, ('SNP')])
```

```
manhattan(Dxy.table, chr = "CHROM", bp="POS", p = "Dxy_100", ylab =
'',
          snp = "SNP",
          suggestiveline = min(dxy.outliers)*100,
          highlight = dxy.outliers.index,
          col = seq(1, 2, by=1), chrlabs = NULL, logp = F)

manhattan(Dxy.table, chr = "CHROM", bp="POS", p = "Fst", ylab = '',
          snp = "SNP",
          suggestiveline = min(fst.outliers),
          highlight = fst.outliers.index,
          col = seq(1, 2, by=1), chrlabs = NULL, logp = F)

lm.dxyfst <- lm(Dxy.table$Dxy_100 ~ Dxy.table$Fst)
summary(lm.dxyfst)
cor(Dxy.table$Dxy_100, Dxy.table$Fst)
```

APPENDIX I: Museum specimens

List of museum specimens used for morphological and colorimetric measurements. AMNH: American Museum of Natural History; LAMNH: Los Angeles Museum of Natural History; MLZ: Moore Laboratory of Zoology at Occidental College; MVZ: Museum of Vertebrate Zoology, University of California Berkeley; NMNH: National Museum of Natural History; SDMNH: San Diego Museum of Natural History.

Collection	Catalog Number	Taxon	Country	State	Sex
AMNH	52894	*dorsalis*	US	AZ	F
AMNH	88570	*pinosus*	US	CA	F
AMNH	88576	*thurberi*	US	CA	F
AMNH	88585	*thurberi*	US	CA	M
AMNH	88597	*thurberi*	US	CA	F
AMNH	88607	*thurberi*	US	CA	M
AMNH	88680	*thurberi*	US	CA	F
AMNH	88693	*thurberi*	US	CA	M
AMNH	88695	*thurberi*	US	CA	F
AMNH	88696	*thurberi*	US	CA	F
AMNH	88707	*pinosus*	US	CA	F
AMNH	88827	*mearnsi*	US	AZ	F
AMNH	114055	*caniceps*	US	CO	F
AMNH	119492	*hyemalis*	CA	AB	F
AMNH	393147	*vulcani*	CR	CR	M
AMNH	393148	*vulcani*	CR	CR	M
AMNH	393151	*vulcani*	CR	CR	F
AMNH	393154	*vulcani*	CR	CR	F
AMNH	393157	*vulcani*	CR	CR	M
AMNH	397940	*alticola*	GU	GUAT	M
AMNH	397941	*alticola*	GU	GUAT	M
AMNH	397942	*alticola*	GU	GUAT	M
AMNH	397943	*alticola*	GU	GUAT	F
AMNH	397945	*alticola*	GU	GUAT	M
AMNH	397946	*alticola*	GU	GUAT	M
AMNH	397947	*alticola*	GU	GUAT	F
AMNH	397948	*alticola*	GU	GUAT	M
AMNH	397953	*alticola*	GU	GUAT	F
AMNH	397954	*alticola*	GU	GUAT	F
AMNH	397955	*alticola*	GU	GUAT	M
AMNH	397956	*alticola*	GU	GUAT	M
AMNH	397957	*alticola*	GU	GUAT	M
AMNH	397958	*alticola*	GU	GUAT	M

AMNH	399317	*alticola*	GU	GUAT	M
AMNH	399318	*alticola*	GU	GUAT	M
AMNH	399319	*alticola*	GU	GUAT	F
AMNH	399320	*alticola*	GU	GUAT	F
AMNH	399321	*alticola*	GU	GUAT	M
AMNH	402265	*hyemalis*	CA	QB	F
AMNH	402782	*pinosus*	US	CA	M
AMNH	402784	*pinosus*	US	CA	M
AMNH	402785	*pinosus*	US	CA	M
AMNH	402789	*pinosus*	US	CA	M
AMNH	402794	*pinosus*	US	CA	F
AMNH	402795	*pinosus*	US	CA	F
AMNH	402796	*pinosus*	US	CA	F
AMNH	402797	*pinosus*	US	CA	F
AMNH	402800	*pinosus*	US	CA	F
AMNH	402825	*townsendi*	MX	BCN	F
AMNH	402828	*townsendi*	MX	BCN	F
AMNH	402829	*townsendi*	MX	BCN	F
AMNH	402831	*townsendi*	MX	BCN	F
AMNH	402832	*townsendi*	MX	BCN	F
AMNH	402875	*bairdi*	MX	BCN	F
AMNH	402877	*bairdi*	MX	BCN	F
AMNH	402904	*mearnsi*	US	AZ	M
AMNH	402907	*mearnsi*	US	AZ	M
AMNH	402908	*mearnsi*	US	AZ	F
AMNH	402912	*mearnsi*	US	AZ	F
AMNH	402913	*mearnsi*	US	AZ	M
AMNH	402954	*caniceps*	US	NM	M
AMNH	402955	*caniceps*	US	NM	M
AMNH	402956	*caniceps*	US	NM	M
AMNH	402957	*caniceps*	US	NM	M
AMNH	402958	*caniceps*	US	NM	M
AMNH	402959	*caniceps*	US	NM	F
AMNH	402960	*caniceps*	US	NM	F
AMNH	402961	*caniceps*	US	NM	M
AMNH	402962	*caniceps*	US	NM	M
AMNH	402964	*caniceps*	US	NM	M
AMNH	402967	*caniceps*	US	NM	M
AMNH	402985	*caniceps*	US	CO	M
AMNH	402988	*caniceps*	US	CO	M
AMNH	403041	*dorsalis*	US	NM	M
AMNH	403045	*dorsalis*	US	NM	F
AMNH	403047	*dorsalis*	US	NM	F
AMNH	403048	*dorsalis*	US	NM	F
AMNH	403049	*dorsalis*	US	NM	F
AMNH	403051	*dorsalis*	US	NM	M

AMNH	442604	*townsendi*	MX	BCN	M
AMNH	461710	*mearnsi*	US	AZ	M
AMNH	518367	*townsendi*	MX	BCN	M
AMNH	762751	*pinosus*	US	CA	F
AMNH	762771	*insularis*	MX	BCN	M
AMNH	762773	*insularis*	MX	BCN	F
AMNH	762774	*insularis*	MX	BCN	F
AMNH	762775	*insularis*	MX	BCN	M
AMNH	762777	*insularis*	MX	BCN	M
AMNH	762783	*insularis*	MX	BCN	F
AMNH	762784	*insularis*	MX	BCN	F
AMNH	762785	*insularis*	MX	BCN	F
AMNH	762786	*insularis*	MX	BCN	M
AMNH	762788	*insularis*	MX	BCN	M
AMNH	762795	*bairdi*	MX	BCN	F
AMNH	762796	*bairdi*	MX	BCN	F
AMNH	762801	*bairdi*	MX	BCN	F
AMNH	762803	*bairdi*	MX	BCN	M
AMNH	762807	*bairdi*	MX	BCN	F
AMNH	762815	*mearnsi*	US	AZ	F
AMNH	762820	*mearnsi*	US	AZ	M
AMNH	762928	*dorsalis*	US	NM	F
AMNH	811944	*vulcani*	CR	CR	M
AMNH	811945	*vulcani*	CR	CR	F
AMNH	811946	*vulcani*	CR	CR	M
AMNH	811947	*vulcani*	CR	CR	F
LAMNH	4938	*aikeni*	US	MT	M
LAMNH	4939	*aikeni*	US	MT	F
LAMNH	9311	*aikeni*	US	CO	M
LAMNH	9312	*aikeni*	US	CO	F
LAMNH	15661	*vulcani*	CR	CR	M
LAMNH	15662	*vulcani*	CR	CR	M
LAMNH	15663	*vulcani*	CR	CR	M
LAMNH	15664	*vulcani*	CR	CR	F
LAMNH	15665	*vulcani*	CR	CR	F
LAMNH	15666	*vulcani*	CR	CR	F
LAMNH	17931	*alticola*	GU	GUAT	M
LAMNH	17932	*alticola*	GU	GUAT	M
LAMNH	17933	*alticola*	GU	GUAT	F
LAMNH	66487	*thurberi*	US	CA	F
LAMNH	66488	*thurberi*	US	CA	M
MLZ	9149	*palliatus*	MX	CHIH	F
MLZ	9151	*palliatus*	MX	CHIH	M
MLZ	9152	*palliatus*	MX	CHIH	F
MLZ	9153	*palliatus*	MX	CHIH	F
MLZ	18462	*palliatus*	MX	CHIH	M

MLZ	18464	*palliatus*	MX	CHIH	F
MLZ	18465	*palliatus*	MX	CHIH	F
MLZ	18727	*palliatus*	MX	CHIH	M
MLZ	18728	*palliatus*	MX	CHIH	M
MLZ	18730	*palliatus*	MX	CHIH	F
MLZ	18733	*palliatus*	MX	CHIH	F
MLZ	18938	*palliatus*	MX	CHIH	M
MLZ	19023	*palliatus*	MX	CHIH	M
MLZ	19025	*palliatus*	MX	CHIH	M
MLZ	32595	*phaeonotus*	MX	MEX	F
MLZ	32598	*phaeonotus*	MX	MEX	M
MLZ	32600	*phaeonotus*	MX	MEX	F
MLZ	32601	*phaeonotus*	MX	MEX	F
MLZ	32618	*phaeonotus*	MX	MEX	M
MLZ	32644	*phaeonotus*	MX	MEX	F
MLZ	32668	*phaeonotus*	MX	MEX	F
MLZ	32671	*phaeonotus*	MX	MEX	M
MLZ	32853	*phaeonotus*	MX	MEX	M
MLZ	32854	*phaeonotus*	MX	MEX	F
MLZ	32870	*phaeonotus*	MX	MEX	F
MLZ	37011	*alticola*	MX	CHIS	M
MLZ	37089	*alticola*	MX	CHIS	M
MLZ	37110	*alticola*	MX	CHIS	F
MLZ	37122	*alticola*	MX	CHIS	M
MLZ	37335	*alticola*	MX	CHIS	M
MLZ	37396	*alticola*	MX	CHIS	F
MLZ	37428	*alticola*	MX	CHIS	M
MLZ	37781	*phaeonotus*	MX	OAX	F
MLZ	37910	*phaeonotus*	MX	OAX	M
MLZ	37911	*phaeonotus*	MX	OAX	F
MLZ	41233	*phaeonotus*	MX	MEX	F
MLZ	41241	*phaeonotus*	MX	MEX	M
MLZ	41242	*phaeonotus*	MX	MEX	F
MLZ	45188	*fulvescens*	MX	CHIS	M
MLZ	45189	*fulvescens*	MX	CHIS	M
MLZ	45190	*fulvescens*	MX	CHIS	M
MLZ	45191	*fulvescens*	MX	CHIS	M
MLZ	45192	*fulvescens*	MX	CHIS	F
MLZ	45193	*fulvescens*	MX	CHIS	M
MLZ	45911	*phaeonotus*	MX	GRO	M
MLZ	46115	*phaeonotus*	MX	GRO	M
MLZ	46126	*phaeonotus*	MX	GRO	F
MLZ	46143	*phaeonotus*	MX	GRO	F
MLZ	46205	*phaeonotus*	MX	GRO	M
MLZ	46267	*phaeonotus*	MX	GRO	M
MLZ	46270	*phaeonotus*	MX	GRO	F

MLZ	46291	*phaeonotus*	MX	GRO	M
MLZ	46301	*phaeonotus*	MX	GRO	F
MLZ	46310	*phaeonotus*	MX	GRO	M
MLZ	46329	*phaeonotus*	MX	GRO	M
MLZ	46331	*phaeonotus*	MX	GRO	M
MLZ	46332	*phaeonotus*	MX	GRO	F
MLZ	56826	*fulvescens*	MX	CHIS	M
MLZ	56836	*fulvescens*	MX	CHIS	F
MLZ	56839	*fulvescens*	MX	CHIS	M
MLZ	56845	*fulvescens*	MX	CHIS	F
MLZ	56882	*fulvescens*	MX	CHIS	F
MLZ	56885	*fulvescens*	MX	CHIS	F
MLZ	56912	*fulvescens*	MX	CHIS	F
MLZ	56913	*fulvescens*	MX	CHIS	M
MLZ	56920	*fulvescens*	MX	CHIS	M
MLZ	56979	*fulvescens*	MX	CHIS	M
MLZ	57047	*fulvescens*	MX	CHIS	M
MLZ	57241	*phaeonotus*	MX	VER	M
MLZ	57315	*phaeonotus*	MX	VER	M
MLZ	58335	*palliatus*	MX	DGO	F
MLZ	58349	*palliatus*	MX	DGO	F
MLZ	58383	*palliatus*	MX	DGO	M
MLZ	65357	*phaeonotus*	MX	OAX	F
MLZ	65358	*phaeonotus*	MX	OAX	M
MLZ	66869	*fulvescens*	MX	CHIS	M
MLZ	66871	*fulvescens*	MX	CHIS	M
MVZ	23850	*thurberi*	US	CA	M
MVZ	23852	*thurberi*	US	CA	F
MVZ	23853	*thurberi*	US	CA	F
MVZ	23854	*thurberi*	US	CA	M
MVZ	23855	*thurberi*	US	CA	F
MVZ	23856	*thurberi*	US	CA	M
MVZ	27745	*dorsalis*	US	AZ	F
MVZ	46336	*townsendi*	MX	BCN	F
MVZ	46337	*townsendi*	MX	BCN	M
MVZ	46339	*townsendi*	MX	BCN	M
MVZ	46340	*townsendi*	MX	BCN	M
MVZ	46341	*townsendi*	MX	BCN	M
MVZ	46342	*townsendi*	MX	BCN	M
MVZ	46343	*townsendi*	MX	BCN	M
MVZ	46344	*townsendi*	MX	BCN	F
MVZ	46345	*townsendi*	MX	BCN	M?
MVZ	47007	*townsendi*	MX	BCN	F
MVZ	47461	*townsendi*	MX	BCN	F
MVZ	47467	*townsendi*	MX	BCN	F
MVZ	48125	*townsendi*	MX	BCN	F

MVZ	48533	*thurberi*	US	CA	M
MVZ	48534	*thurberi*	US	CA	M
MVZ	48535	*thurberi*	US	CA	M
MVZ	48537	*thurberi*	US	CA	M
MVZ	48538	*thurberi*	US	CA	F
MVZ	55520	*bairdi*	MX	BCS	M
MVZ	55521	*bairdi*	MX	BCS	F
MVZ	55522	*bairdi*	MX	BCS	M
MVZ	55523	*bairdi*	MX	BCS	M
MVZ	55524	*bairdi*	MX	BCS	M
MVZ	55525	*bairdi*	MX	BCS	F
MVZ	55526	*bairdi*	MX	BCS	M
MVZ	55527	*bairdi*	MX	BCS	M
MVZ	55528	*bairdi*	MX	BCS	M
MVZ	55529	*bairdi*	MX	BCS	F
MVZ	55530	*bairdi*	MX	BCS	M
MVZ	55531	*bairdi*	MX	BCS	M
MVZ	55532	*bairdi*	MX	BCS	F
MVZ	55533	*bairdi*	MX	BCS	F
MVZ	55534	*bairdi*	MX	BCS	M
MVZ	55540	*bairdi*	MX	BCS	M
MVZ	55547	*bairdi*	MX	BCS	M
MVZ	58182	*dorsalis*	US	AZ	M
MVZ	58183	*dorsalis*	US	AZ	M
MVZ	58197	*dorsalis*	US	AZ	M
MVZ	58200	*dorsalis*	US	AZ	M
MVZ	58201	*dorsalis*	US	AZ	M
MVZ	58202	*dorsalis*	US	AZ	M
MVZ	58203	*dorsalis*	US	AZ	F
MVZ	58204	*dorsalis*	US	AZ	M
MVZ	59041	*dorsalis*	US	NM	F?
MVZ	60911	*hyemalis*	CA	AB	M
MVZ	60913	*hyemalis*	CA	AB	F
MVZ	60914	*hyemalis*	CA	AB	F
MVZ	60918	*hyemalis*	CA	AB	M
MVZ	60919	*hyemalis*	CA	AB	M
MVZ	60920	*hyemalis*	CA	AB	F
MVZ	60921	*hyemalis*	CA	AB	M
MVZ	62288	*thurberi*	US	CA	M
MVZ	62289	*thurberi*	US	CA	F
MVZ	62290	*thurberi*	US	CA	M
MVZ	63018	*thurberi*	US	CA	M
MVZ	63019	*thurberi*	US	CA	F
MVZ	65697	*montanus*	CA	BC	F
MVZ	65698	*montanus*	CA	BC	M
MVZ	65699	*montanus*	CA	BC	F

MVZ	65700	*montanus*	CA	BC	M
MVZ	65701	*montanus*	CA	BC	M
MVZ	65703	*montanus*	CA	BC	M
MVZ	65705	*montanus*	CA	BC	M
MVZ	73241	*pinosus*	US	CA	M
MVZ	73242	*pinosus*	US	CA	M
MVZ	73243	*pinosus*	US	CA	F
MVZ	73244	*pinosus*	US	CA	M
MVZ	73245	*pinosus*	US	CA	M
MVZ	73246	*pinosus*	US	CA	M
MVZ	73248	*pinosus*	US	CA	F
MVZ	79982	*aikeni*	US	MT	M
MVZ	80605	*alticola*	MX	CHIS	M
MVZ	83086	*dorsalis*	US	AZ	F
MVZ	87312	*pinosus*	US	CA	M
MVZ	94643	*montanus*	CA	BC	M
MVZ	94644	*montanus*	CA	BC	M
MVZ	94645	*montanus*	CA	BC	F
MVZ	98927	*aikeni*	US	MT	M
MVZ	98928	*aikeni*	US	MT	F
MVZ	98929	*aikeni*	US	MT	F
MVZ	98945	*aikeni*	US	MT	F
MVZ	98946	*aikeni*	US	MT	F
MVZ	98949	*aikeni*	US	MT	M
MVZ	98950	*aikeni*	US	MT	M
MVZ	98951	*aikeni*	US	MT	M
MVZ	98952	*aikeni*	US	MT	M
MVZ	98953	*aikeni*	US	MT	M
MVZ	98954	*aikeni*	US	MT	F
MVZ	98959	*mearnsi*	US	MT	F
MVZ	98960	*mearnsi*	US	MT	M
MVZ	98961	*mearnsi*	US	MT	M
MVZ	98963	*mearnsi*	US	MT	F
MVZ	98965	*mearnsi*	US	MT	M
MVZ	98967	*mearnsi*	US	MT	M
MVZ	98968	*mearnsi*	US	MT	F
MVZ	98970	*mearnsi*	US	MT	M
MVZ	98971	*mearnsi*	US	MT	M
MVZ	98972	*mearnsi*	US	MT	F
MVZ	98973	*mearnsi*	US	MT	M
MVZ	98974	*mearnsi*	US	MT	M
MVZ	98975	*mearnsi*	US	MT	M
MVZ	98976	*mearnsi*	US	MT	F
MVZ	98977	*mearnsi*	US	MT	F
MVZ	98979	*mearnsi*	US	MT	M
MVZ	98981	*mearnsi*	US	MT	M

MVZ	98982	*mearnsi*	US	MT	F
MVZ	98983	*mearnsi*	US	MT	F
MVZ	98984	*mearnsi*	US	MT	M
MVZ	98985	*mearnsi*	US	MT	F
MVZ	98986	*mearnsi*	US	MT	M
MVZ	98987	*mearnsi*	US	MT	F
MVZ	106175	*dorsalis*	US	NM	F
MVZ	106176	*dorsalis*	US	NM	F
MVZ	106178	*dorsalis*	US	AZ	M
MVZ	121543	*fulvescens*	MX	CHIS	M
MVZ	121544	*fulvescens*	MX	CHIS	F
MVZ	121545	*fulvescens*	MX	CHIS	M
MVZ	121546	*fulvescens*	MX	CHIS	M
MVZ	121547	*fulvescens*	MX	CHIS	M
MVZ	129425	*hyemalis*	US	AK	M
MVZ	133330	*caniceps*	US	NV	M
MVZ	133331	*caniceps*	US	NV	F
MVZ	133332	*caniceps*	US	NV	M
MVZ	133334	*caniceps*	US	NV	F
MVZ	133335	*caniceps*	US	NV	M
MVZ	133336	*caniceps*	US	NV	F
MVZ	133338	*caniceps*	US	NV	F
MVZ	133343	*caniceps*	US	NV	F
MVZ	133344	*caniceps*	US	NV	F
MVZ	137476	*hyemalis*	US	AK	M
MVZ	137478	*hyemalis*	US	AK	M
MVZ	140262	*aikeni*	US	SD	M
MVZ	141646	*hyemalis*	US	AK	M
MVZ	141647	*hyemalis*	US	AK	M
MVZ	141648	*hyemalis*	US	AK	M
MVZ	141649	*hyemalis*	US	AK	F
MVZ	141650	*hyemalis*	US	AK	M
MVZ	141651	*hyemalis*	US	AK	F
MVZ	141652	*hyemalis*	US	AK	M
MVZ	141653	*hyemalis*	US	AK	F
MVZ	141654	*hyemalis*	US	AK	F
MVZ	141655	*hyemalis*	US	AK	F
MVZ	155375	*fulvescens*	MX	CHIS	M
MVZ	155376	*fulvescens*	MX	CHIS	M
MVZ	158636	*hyemalis*	US	AK	M
MVZ	158638	*hyemalis*	US	AK	M
MVZ	177269	*pinosus*	US	CA	M
MVZ	177270	*pinosus*	US	CA	M
MVZ	177271	*pinosus*	US	CA	F
MVZ	177273	*pinosus*	US	CA	M
MVZ	177275	*pinosus*	US	CA	F

MVZ	177278	*pinosus*	US	CA	M
MVZ	177279	*pinosus*	US	CA	M
MVZ	177280	*pinosus*	US	CA	M
MVZ	181959	*pinosus*	US	CA	F
MVZ	183129	*pinosus*	US	CA	M
NMNH	66902	*caniceps*	US	CO	M
NMNH	87790	*vulcani*	CR	CR	F
NMNH	133545	*thurberi*	US	CA	M
NMNH	133546	*thurberi*	US	CA	M
NMNH	133551	*thurberi*	US	CA	F
NMNH	133553	*thurberi*	US	CA	M
NMNH	133554	*thurberi*	US	CA	M
NMNH	133555	*thurberi*	US	CA	F
NMNH	133556	*thurberi*	US	CA	M
NMNH	133856	*thurberi*	US	CA	M
NMNH	133857	*thurberi*	US	CA	F
NMNH	133858	*thurberi*	US	CA	F
NMNH	134269	*thurberi*	US	CA	F
NMNH	134270	*thurberi*	US	CA	F
NMNH	134272	*thurberi*	US	CA	M
NMNH	134273	*thurberi*	US	CA	F
NMNH	134275	*thurberi*	US	CA	M
NMNH	138509	*dorsalis*	US	AZ	M
NMNH	138544	*mearnsi*	US	WY	M
NMNH	138578	*oreganus*	US	AK	F
NMNH	143899	*fulvescens*	MX	CHIS	F
NMNH	143900	*fulvescens*	MX	CHIS	F
NMNH	143904	*fulvescens*	MX	CHIS	M
NMNH	143908	*fulvescens*	MX	CHIS	M
NMNH	143910	*fulvescens*	MX	CHIS	F
NMNH	143913	*fulvescens*	MX	CHIS	M
NMNH	143998	*alticola*	GU	GUAT	M
NMNH	143999	*alticola*	GU	GUAT	F
NMNH	144000	*alticola*	GU	GUAT	F
NMNH	144002	*alticola*	GU	GUAT	F
NMNH	144004	*alticola*	GU	GUAT	F
NMNH	144006	*alticola*	GU	GUAT	F
NMNH	144007	*alticola*	GU	GUAT	F
NMNH	144008	*alticola*	GU	GUAT	F
NMNH	144009	*alticola*	GU	GUAT	F
NMNH	144012	*alticola*	GU	GUAT	M
NMNH	144013	*alticola*	GU	GUAT	F
NMNH	149682	*hyemalis*	US	NH	M
NMNH	157701	*oreganus*	CA	BC	F
NMNH	157808	*hyemalis*	CA	AB	M
NMNH	158395	*caniceps*	US	NV	M

NMNH	158396	*caniceps*	US	NV	M
NMNH	163944	*oreganus*	CA	BC	F
NMNH	164921	*oreganus*	US	AK	M
NMNH	164922	*oreganus*	US	AK	F
NMNH	164924	*oreganus*	US	AK	M
NMNH	165163	*oreganus*	US	AK	M
NMNH	165410	*carolinensis*	US	MD	M
NMNH	166843	*oreganus*	CA	BC	F
NMNH	170234	*oreganus*	US	AK	M
NMNH	183205	*hyemalis*	CA	AB	F
NMNH	183207	*hyemalis*	CA	AB	F
NMNH	186190	*oreganus*	US	AK	M
NMNH	186244	*oreganus*	US	AK	F
NMNH	186245	*oreganus*	US	AK	M
NMNH	186246	*oreganus*	US	AK	M
NMNH	186247	*oreganus*	US	AK	M
NMNH	186248	*oreganus*	US	AK	M
NMNH	186249	*oreganus*	US	AK	F
NMNH	186659	*oreganus*	US	AK	M
NMNH	186660	*oreganus*	US	AK	F
NMNH	192969	*hyemalis*	CA	NT	F
NMNH	194248	*fulvescens*	MX	CHIS	F
NMNH	194249	*fulvescens*	MX	CHIS	M
NMNH	194250	*fulvescens*	MX	CHIS	F
NMNH	194847	*hyemalis*	CA	NT	F
NMNH	196965	*pontilis*	MX	BCN	M
NMNH	196997	*dorsalis*	US	NM	M
NMNH	197003	*hyemalis*	CA	NT	F
NMNH	197005	*caniceps*	US	CO	M
NMNH	199479	*vulcani*	CR	CR	M
NMNH	199483	*vulcani*	CR	CR	M
NMNH	199484	*vulcani*	CR	CR	M
NMNH	199485	*vulcani*	CR	CR	M
NMNH	199486	*vulcani*	CR	CR	F
NMNH	199490	*vulcani*	CR	CR	F
NMNH	199492	*vulcani*	CR	CR	M
NMNH	199493	*vulcani*	CR	CR	F
NMNH	200138	*vulcani*	CR	CR	M
NMNH	200139	*vulcani*	CR	CR	F
NMNH	228420	*mearnsi*	US	WY	M
NMNH	228421	*mearnsi*	US	WY	M
NMNH	239638	*hyemalis*	US	AK	F
NMNH	268391	*mearnsi*	US	MT	F
NMNH	268392	*mearnsi*	US	MT	F
NMNH	268562	*mearnsi*	US	MT	F
NMNH	268563	*mearnsi*	US	MT	M

NMNH	268564	*mearnsi*	US	MT	F
NMNH	268635	*mearnsi*	US	MT	M
NMNH	268636	*mearnsi*	US	MT	F
NMNH	268637	*mearnsi*	US	MT	M
NMNH	271183	*oreganus*	US	AK	F
NMNH	286561	*oreganus*	US	AK	M
NMNH	301819	*carolinensis*	US	NC	M
NMNH	301830	*carolinensis*	US	NC	F
NMNH	337681	*carolinensis*	US	KY	F
NMNH	337683	*carolinensis*	US	KY	M
NMNH	338221	*carolinensis*	US	GA	M
NMNH	348231	*carolinensis*	US	VA	M
NMNH	348232	*carolinensis*	US	VA	M
NMNH	348835	*carolinensis*	US	WV	M
NMNH	350791	*carolinensis*	US	VA	F
NMNH	350795	*carolinensis*	US	VA	F
NMNH	351546	*carolinensis*	US	TN	M
NMNH	351548	*carolinensis*	US	TN	F
NMNH	351549	*carolinensis*	US	TN	F
NMNH	351553	*carolinensis*	US	TN	M
NMNH	351666	*hyemalis*	CA	NS	M
NMNH	351667	*hyemalis*	CA	NS	M
NMNH	351668	*hyemalis*	CA	NS	F
NMNH	357371	*carolinensis*	US	VA	M
NMNH	357985	*carolinensis*	US	NC	F
NMNH	363310	*carolinensis*	US	GA	F
NMNH	363653	*oreganus*	US	WA	M
NMNH	379431	*carolinensis*	US	GA	M
NMNH	381876	*hyemalis*	CA	NL	M
NMNH	381877	*hyemalis*	CA	NL	F
NMNH	381878	*hyemalis*	CA	NL	F
NMNH	394025	*hyemalis*	CA	NL	M
NMNH	394027	*hyemalis*	CA	NL	M
NMNH	394032	*hyemalis*	CA	NL	F
NMNH	394306	*carolinensis*	US	WV	M
NMNH	397093	*hyemalis*	US	ME	M
NMNH	397454	*mearnsi*	US	ID	F
NMNH	407352	*caniceps*	US	CO	M
NMNH	418382	*mearnsi*	US	ID	F
NMNH	442631	*oreganus*	US	OR	F
NMNH	464080	*mearnsi*	US	ID	F
NMNH	466133	*mearnsi*	US	MT	M
NMNH	478975	*pinosus*	US	CA	F
NMNH	480046	*pinosus*	US	CA	F
NMNH	481974	*mearnsi*	US	WY	M
NMNH	486209	*vulcani*	CR	CR	M

NMNH	486210	*vulcani*	CR	CR	F
NMNH	486212	*vulcani*	CR	CR	F
NMNH	486213	*vulcani*	CR	CR	F
NMNH	486218	*vulcani*	CR	CR	F
NMNH	486219	*vulcani*	CR	CR	M
NMNH	486222	*vulcani*	CR	CR	M
NMNH	486223	*vulcani*	CR	CR	M
NMNH	486224	*vulcani*	CR	CR	F
NMNH	486225	*vulcani*	CR	CR	M
NMNH	529945	*hyemalis*	US	AK	M
NMNH	529946	*hyemalis*	US	AK	F
NMNH	529950	*hyemalis*	US	AK	M
NMNH	532011	*pinosus*	US	CA	M
NMNH	563360	*pinosus*	US	CA	M
NMNH	563365	*pinosus*	US	CA	F
NMNH	563374	*pinosus*	US	CA	F
NMNH	563381	*pinosus*	US	CA	M
NMNH	566480	*pinosus*	US	CA	M
NMNH	566481	*pinosus*	US	CA	M
NMNH	566766	*pinosus*	US	CA	M
NMNH	566767	*pinosus*	US	CA	M
NMNH	566770	*pinosus*	US	CA	M
NMNH	566788	*pinosus*	US	CA	M
NMNH	566791	*pinosus*	US	CA	F
NMNH	566792	*pinosus*	US	CA	F
NMNH	566794	*pinosus*	US	CA	F
NMNH	566801	*pinosus*	US	CA	F
NMNH	566804	*pinosus*	US	CA	F
NMNH	566809	*pinosus*	US	CA	F
NMNH	594692	*hyemalis*	CA	QB	M
NMNH	594696	*oreganus*	US	AK	F
NMNH	594743	*caniceps*	US	CO	F
NMNH	594780	*carolinensis*	US	NC	F
NMNH	594831	*caniceps*	US	UT	F
NMNH	601455	*carolinensis*	US	VA	F
NMNH	601503	*carolinensis*	US	VA	F
SDMNH	1025	*aikeni*	US	CO	M
SDMNH	1026	*aikeni*	US	CO	F
SDMNH	1056	*palliatus*	US	AZ	M
SDMNH	8744	*townsendi*	MX	BCN	M
SDMNH	8749	*townsendi*	MX	BCN	M
SDMNH	8750	*townsendi*	MX	BCN	M
SDMNH	8764	*townsendi*	MX	BCN	M
SDMNH	8777	*townsendi*	MX	BCN	M
SDMNH	8778	*townsendi*	MX	BCN	F
SDMNH	8793	*townsendi*	MX	BCN	F

SDMNH	8794	*townsendi*	MX	BCN	F
SDMNH	8795	*townsendi*	MX	BCN	M
SDMNH	8808	*townsendi*	MX	BCN	M
SDMNH	8856	*townsendi*	MX	BCN	M
SDMNH	9413	*pontilis*	MX	BCN	F
SDMNH	9414	*pontilis*	MX	BCN	F
SDMNH	9429	*pontilis*	MX	BCN	M
SDMNH	10828	*townsendi*	MX	BCN	M
SDMNH	10846	*townsendi*	MX	BCN	M
SDMNH	10887	*pontilis*	MX	BCN	M
SDMNH	12631	*insularis*	MX	BCN	M
SDMNH	12632	*insularis*	MX	BCN	M
SDMNH	12634	*insularis*	MX	BCN	M
SDMNH	12635	*insularis*	MX	BCN	M
SDMNH	12637	*insularis*	MX	BCN	M
SDMNH	12640	*insularis*	MX	BCN	M
SDMNH	14653	*palliatus*	US	AZ	F
SDMNH	14687	*palliatus*	US	AZ	M
SDMNH	14809	*palliatus*	US	AZ	F
SDMNH	14810	*palliatus*	US	AZ	M
SDMNH	14812	*palliatus*	US	AZ	M
SDMNH	14813	*palliatus*	US	AZ	M
SDMNH	14922	*insularis*	MX	BCN	F
SDMNH	14923	*insularis*	MX	BCN	M
SDMNH	14924	*insularis*	MX	BCN	M
SDMNH	14925	*insularis*	MX	BCN	M
SDMNH	14926	*insularis*	MX	BCN	F
SDMNH	14927	*insularis*	MX	BCN	F
SDMNH	14928	*insularis*	MX	BCN	F
SDMNH	14929	*insularis*	MX	BCN	F
SDMNH	14932	*insularis*	MX	BCN	F
SDMNH	14950	*insularis*	MX	BCN	F
SDMNH	14951	*insularis*	MX	BCN	M
SDMNH	15770	*palliatus*	US	AZ	M
SDMNH	15771	*palliatus*	US	AZ	F
SDMNH	15772	*palliatus*	US	AZ	F
SDMNH	15776	*palliatus*	US	AZ	F
SDMNH	15837	*palliatus*	US	AZ	F
SDMNH	16256	*dorsalis*	US	AZ	M
SDMNH	16278	*dorsalis*	US	AZ	M
SDMNH	16284	*dorsalis*	US	AZ	M
SDMNH	17022	*montanus*	US	CA	F
SDMNH	17236	*pontilis*	MX	BCN	F
SDMNH	18580	*montanus*	US	AZ	F
SDMNH	28412	*aikeni*	US	CO	M
SDMNH	28413	*aikeni*	US	CO	F

SDMNH	28414	*aikeni*	US	CO	F
SDMNH	28416	*aikeni*	US	CO	F
SDMNH	28498	*montanus*	US	OR	M
SDMNH	28499	*montanus*	US	OR	M
SDMNH	28500	*montanus*	US	OR	M
SDMNH	28501	*montanus*	US	OR	M
SDMNH	28502	*montanus*	US	OR	F
SDMNH	28504	*montanus*	US	OR	M
SDMNH	28505	*montanus*	US	OR	M
SDMNH	28509	*montanus*	US	OR	M
SDMNH	28515	*montanus*	US	OR	F
SDMNH	28525	*montanus*	US	OR	F
SDMNH	28547	*montanus*	US	OR	F
SDMNH	28586	*montanus*	US	OR	M
SDMNH	28598	*montanus*	US	OR	F
SDMNH	28675	*townsendi*	MX	BCN	M
SDMNH	28676	*townsendi*	MX	BCN	F
SDMNH	28677	*townsendi*	MX	BCN	F
SDMNH	28692	*insularis*	MX	BCN	M
SDMNH	28693	*bairdi*	MX	BCS	M
SDMNH	28694	*bairdi*	MX	BCS	F
SDMNH	28695	*bairdi*	MX	BCS	F
SDMNH	28696	*bairdi*	MX	BCS	M
SDMNH	34458	*pontilis*	MX	BCN	M
SDMNH	34459	*pontilis*	MX	BCN	M
SDMNH	34460	*pontilis*	MX	BCN	M
SDMNH	34461	*pontilis*	MX	BCN	F
SDMNH	34467	*pontilis*	MX	BCN	F
SDMNH	34473	*pontilis*	MX	BCN	M
SDMNH	34474	*pontilis*	MX	BCN	M
SDMNH	34476	*insularis*	MX	BCN	M
SDMNH	34477	*insularis*	MX	BCN	M
SDMNH	34483	*dorsalis*	US	AZ	F
SDMNH	34484	*dorsalis*	US	AZ	F
SDMNH	34485	*dorsalis*	US	AZ	F?
SDMNH	34489	*palliatus*	US	AZ	F
SDMNH	34490	*palliatus*	US	AZ	F
SDMNH	34495	*palliatus*	US	AZ	M
SDMNH	34496	*palliatus*	US	AZ	M
SDMNH	34497	*palliatus*	US	AZ	M
SDMNH	34498	*palliatus*	US	AZ	F
SDMNH	34499	*palliatus*	US	AZ	M
SDMNH	34501	*palliatus*	US	AZ	F
SDMNH	34517	*bairdi*	MX	BCS	M
SDMNH	34518	*bairdi*	MX	BCS	M
SDMNH	35378	*bairdi*	MX	BCS	M

SDMNH	35379	*bairdi*	MX	BCS	M
SDMNH	35380	*bairdi*	MX	BCS	F
SDMNH	35382	*bairdi*	MX	BCS	M
SDMNH	35385	*bairdi*	MX	BCS	F
SDMNH	35386	*bairdi*	MX	BCS	M
SDMNH	43930	*aikeni*	CA	NB	M
SDMNH	44244	*aikeni*	CA	NB	F

APPENDIX II: GBS samples

List of samples sequenced by genotyping-by-sequencing (GBS). State abbreviations are the following: Virginia (VA), Wyoming (WY), Alaska (AK), California (CA), Colorado (CO), Idaho (ID), Utah (UT), Maine (ME), New Hampshire (NH), New Mexico (NM), Nevada (NV), Oregon (OR), Utah (UT), Wyoming (WY), and Arizona (AZ) in the USA; British Columbia (BC) in Canada; Baja California Norte (BC N), Baja California Sur (BC S), Chihuahua (CHIH), Coahuila (COAH), Mexico City (CM), Durango (DGO), Michoacán (MICH), Nuevo León (NL), Oaxaca (OAX), and Chiapas (CHIS) in Mexico; and Huehuetenango (HUE) in Guatemala.

Taxonomy	Plate	Locality	State	Country	LAT	LONG	Collection	SEX
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-104.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-104.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-104.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	1	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco phaeonotus alticola	1	Culchemal	HUE	Guatemala			5/18/2001	M
Junco phaeonotus alticola	1	Culchemal	HUE	Guatemala			5/18/2001	M
Junco phaeonotus alticola	1	Culchemal	HUE	Guatemala			5/18/2001	U
Junco phaeonotus alticola	1	Chichim	HUE	Guatemala			5/19/2001	M
Junco phaeonotus alticola	1	Chichim	HUE	Guatemala			5/19/2001	M
Junco phaeonotus alticola	1	Chichim	HUE	Guatemala			5/20/2001	M
Junco phaeonotus alticola	1	Chichim	HUE	Guatemala			5/26/2003	U
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/10/2011	U
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/10/2011	U
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/10/2011	U
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/10/2011	U
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/11/2011	M
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/11/2011	M
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/11/2011	U
Junco hyemalis caniceps	1	Uinta Range	UT	USA	40.7	-110.9	7/16/2005	M
Junco hyemalis caniceps	1	Uinta Range	UT	USA	40.7	-110.9	7/16/2005	M
Junco hyemalis caniceps	1	Uinta Range	UT	USA	40.7	-110.9	7/16/2005	M
Junco hyemalis caniceps	1	Uinta Range	UT	USA	40.7	-111.0	7/16/2005	M
Junco hyemalis caniceps	1	Uinta Range	UT	USA	40.7	-111.0	7/16/2005	M
Junco hyemalis caniceps	1	Dixie NF	UT	USA	37.5	-112.5	7/17/2005	M
Junco hyemalis caniceps	1	Dixie NF	UT	USA	37.5	-112.5	7/17/2005	M
Junco hyemalis caniceps	1	Dixie NF	UT	USA	37.5	-112.6	7/18/2005	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/20/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/20/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/20/2012	M
Junco hyemalis carolinensis	1	Mountain Lake	VA	USA			4/20/2012	M
Junco hyemalis dorsalis	1	Big Lake	AZ	USA			6/2/2006	M
Junco hyemalis dorsalis	1	Big Lake	AZ	USA			6/2/2006	M
Junco hyemalis dorsalis	1	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	1	Big Lake	AZ	USA			6/2/2006	M
Junco hyemalis dorsalis	1	Flagstaff	AZ	USA	35.3	-111.6	6/4/2006	M

Junco hyemalis dorsalis	1	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	1	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	1	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco phaeonotus fulvescens	1	Encuentro	CHIS	Mexico			6/11/2010	M
Junco phaeonotus fulvescens	1	Bandidos	CHIS	Mexico	16.7	-92.6	6/30/2006	M
Junco phaeonotus fulvescens	1	Encuentro	CHIS	Mexico	16.7	-92.6	6/30/2006	F
Junco bairdi	1	Sierra de la Laguna	BC S	Mexico			4/10/2011	M
Junco phaeonotus alticola	1	Chichim	HUE	Guatemala			6/9/2010	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco insularis	1	Guadalupe	BC N	Mexico	29.1	-118.3	4/20/2011	M
Junco insularis	1	Estacion GECI	BC N	Mexico	29.1	-118.3	6/18/2010	M
Junco insularis	1	Guadalupe	BC N	Mexico			4/21/2011	M
Junco insularis	1	Guadalupe	BC N	Mexico			4/22/2011	M
Junco insularis	1	Estacion GECI	BC N	Mexico	29.1	-118.3	6/20/2010	M
Junco insularis	1	Guadalupe	BC N	Mexico			4/24/2011	M
Junco insularis	1	Guadalupe	BC N	Mexico			4/25/2011	M
Junco hyemalis mearnsi	1	Shoshone NF	WY	USA			7/11/2005	M
Junco hyemalis mearnsi	1	Shoshone NF	WY	USA			7/11/2005	M
Junco hyemalis mearnsi	1	Shoshone NF	WY	USA			7/12/2005	M
Junco hyemalis mearnsi	1	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	1	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	1	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	1	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.4	6/16/2013	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA			6/16/2013	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco phaeonotus palliatus	1	Bajio de la Vibora	DGO	Mexico			6/2/2002	M
Junco phaeonotus palliatus	1	Bajio de la Vibora	DGO	Mexico			6/2/2002	M
Junco phaeonotus palliatus	1	Sierra Madre Occ.	DGO	Mexico	23.9	-105.2	5/26/2004	M
Junco hyemalis montanus	1	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco phaeonotus palliatus	1	Pinaleno Mts.	AZ	USA	32.7	-109.9	5/31/2006	M
Junco phaeonotus palliatus	1	Bajio de la Vibora	DGO	Mexico			4/24/2001	M
Junco phaeonotus palliatus	1	Bajio de la Vibora	DGO	Mexico			6/2/2002	M
Junco phaeonotus palliatus	1	Pinaleno Mts.	AZ	USA	32.7	-109.9	5/31/2006	M
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	F
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	M
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	M
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	U
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	M
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/23/2000	U
Junco phaeonotus phaeonotus	1	La Cima	DF	Mexico			12/26/2000	U
Junco phaeonotus palliatus	1	La Cima	DF	Mexico			5/10/2001	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.8	-116.4	5/11/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.9	-116.5	5/11/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.9	-116.5	5/11/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.9	-116.5	5/11/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco hyemalis thurberi	1	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco vulcani	1			Costa Rica				
Junco vulcani	1			Costa Rica				
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M

Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	2	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico			5/23/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/24/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.1	-115.9	5/24/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/24/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/24/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/24/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.1	-115.9	5/25/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis pontilis	2	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.8	-116.4	5/10/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.9	-116.5	5/11/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.9	-116.5	5/11/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.8	-116.5	5/11/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.8	-116.4	5/12/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.9	-116.5	5/12/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.9	-116.5	5/12/2012	M
Junco hyemalis thurberi	2	Mt. Laguna	CA	USA	32.8	-116.4	5/13/2012	M
Junco hyemalis pinosus	2	Santa Cruz Mts.	CA	USA	37.0	-122.1	7/4/2014	M
Junco hyemalis pinosus	2	Santa Cruz Mts.	CA	USA	37.0	-122.1	7/4/2014	M
Junco hyemalis pinosus	2	Santa Cruz Mts.	CA	USA			7/4/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/5/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/5/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/5/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/5/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.1	-121.5	7/6/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/6/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/6/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/6/2014	M
Junco hyemalis pinosus	2	Nascimiento Road	CA	USA	36.0	-121.4	7/6/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	2	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.4	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.3	-117.1	6/16/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.2	-117.1	6/17/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.2	-117.0	6/18/2013	M
Junco hyemalis montanus	2	Wallowa NF	OR	USA	45.2	-117.0	6/18/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.4	-121.9	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.4	-121.9	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M

Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/13/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/14/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/14/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/14/2013	M
Junco hyemalis shufeldti	2	Willamette NF	OR	USA	44.5	-122.0	6/14/2013	M
Junco hyemalis oreganus	2	N. Banks Island	BC	Canada			9/2/2004	F
Junco hyemalis oreganus	2	N. Banks Island	BC	Canada			9/2/2004	U
Junco hyemalis oreganus	2	N. Banks Island	BC	Canada			9/2/2004	U
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	U
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	M
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	M
Junco hyemalis oreganus	2	Porcher Island	BC	Canada			9/3/2004	F
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-104.5	7/13/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-104.5	7/13/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.6	-104.4	7/13/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.6	-104.4	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.6	-104.4	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-104.5	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis aikeni	3	Black Hills NF	WY	USA	44.5	-107.5	7/14/2005	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/15/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/15/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA			6/15/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/15/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA			6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA			6/16/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	39.3	-106.7	6/17/2014	M
Junco hyemalis caniceps	3	Aspen	UT	USA	40.4	-111.6	6/12/2006	M
Junco hyemalis caniceps	3	Aspen	UT	USA	40.4	-111.6	6/12/2006	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/15/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/15/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/15/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/16/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/16/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.8	-110.9	7/16/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.7	-111.0	7/16/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.7	-111.0	7/16/2005	M
Junco hyemalis caniceps	3	Uinta Range	UT	USA	40.7	-111.0	7/16/2005	M
Junco hyemalis caniceps	3	Wasatch	UT	USA	40.6	-111.6	6/19/2013	M
Junco hyemalis caniceps	3	Wasatch	UT	USA	40.6	-111.6	6/20/2013	M

Junco hyemalis caniceps	3	Wasatch	UT	USA	40.6	-111.6	6/20/2013	M
Junco hyemalis caniceps	3	Wasatch	UT	USA	40.6	-111.6	6/20/2013	M
Junco hyemalis caniceps	3	Wasatch	UT	USA	40.6	-111.6	6/20/2013	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/13/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/17/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/17/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/19/2012	M
Junco hyemalis carolinensis	3	Mountain Lake	VA	USA			4/20/2012	M
Junco hyemalis hyemalis	3	Rangeley	ME	USA			7/7/2003	M
Junco hyemalis hyemalis	3	Rangeley	ME	USA			7/9/2003	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.3	-71.2	7/12/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.7	7/13/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.2	7/14/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.3	-71.2	7/14/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.3	-71.2	7/15/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.5	7/16/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.5	7/16/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.5	7/16/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.2	-71.5	7/17/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA			7/17/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.3	-71.5	7/17/2014	M
Junco hyemalis hyemalis	3	White Mts.	NH	USA	44.3	-71.5	7/17/2014	M
Junco phaeonotus x dorsalis	3	Big Burro Mts.	NM	USA	32.6	-108.4	6/23/2014	M
Junco phaeonotus x dorsalis	3	Big Burro Mts.	NM	USA	32.6	-108.4	6/24/2014	M
Junco hyemalis mearnsi	3	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	3	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	3	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	3	Bighorn	WY	USA	44.2	-107.2	7/13/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/11/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/11/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/12/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/12/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/12/2005	M
Junco hyemalis mearnsi	3	Lander	WY	USA			7/12/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.5	7/17/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.6	7/17/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.6	7/17/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.6	7/18/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.6	7/18/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.6	7/18/2005	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.8	6/10/2006	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.8	6/10/2006	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.8	6/11/2006	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M

Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M
Junco hyemalis caniceps	4	Rico	CO	USA	37.8	-107.9	6/13/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.3	-105.7	6/18/2014	M
Junco hyemalis caniceps	4	Dixie NF	UT	USA	37.5	-112.5	7/17/2005	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.3	-105.7	6/18/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.3	-105.7	6/18/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA			6/18/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.3	-105.7	6/19/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.1	-105.5	6/19/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.1	-105.5	6/19/2014	M
J. h. caniceps x dorsalis	4	Kaibab Plateau	AZ	USA	36.6	-112.2	6/8/2006	M
J. h. caniceps x dorsalis	4	Kaibab Plateau	AZ	USA	36.6	-112.2	6/9/2006	M
J. h. caniceps x dorsalis	4	Kaibab Plateau	AZ	USA	36.6	-112.2	6/7/2006	M
J. h. caniceps x dorsalis	4	Kaibab Plateau	AZ	USA	36.6	-112.2	6/7/2006	M
J. h. caniceps x dorsalis	4	Kaibab Plateau	AZ	USA	36.6	-112.2	6/7/2006	M
Junco hyemalis caniceps	4	Rico	CO	USA			6/12/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/20/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/20/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/20/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
Junco hyemalis caniceps	4	Sangre de Cristo	CO	USA	38.2	-105.7	6/18/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/22/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/22/2014	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/20/2014	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/5/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/6/2006	M
Junco hyemalis dorsalis	4	Flagstaff	AZ	USA	35.3	-111.6	6/6/2006	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	34.0	-109.6	7/20/2005	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	33.9	-109.5	7/21/2005	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	33.9	-109.5	7/21/2005	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	33.9	-109.5	7/21/2005	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	33.9	-109.5	7/21/2005	U
Junco hyemalis dorsalis	4	Greer	AZ	USA			6/10/2014	M
J. h. caniceps x dorsalis	4	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	34.0	-109.6	6/10/2014	M
Junco hyemalis dorsalis	4	Greer	AZ	USA	34.0	-109.5	6/10/2014	M
Junco hyemalis dorsalis	4	Greer	AZ	USA			6/11/2014	M
Junco hyemalis dorsalis	4	Greer	AZ	USA			6/11/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA			6/25/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA			6/25/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA			6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA			6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M

Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/27/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/27/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/27/2014	M
Junco hyemalis dorsalis	4	Sacramento Mts.	NM	USA	32.9	-105.8	6/27/2014	M
Junco h. thurberi x caniceps	5	Big Creek Canyon	NV	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Grapevine Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Crooked Creek	CA	USA				
Junco h. thurberi x caniceps	5	Grapevine Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Grapevine Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Grapevine Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Grapevine Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Mamooth Lake	CA	USA	37.7	-119.0	5/6/2012	M
Junco h. thurberi x caniceps	5	Mamooth Lake	CA	USA	37.7	-119.0	5/6/2012	M
Junco h. thurberi x caniceps	5	Pamint Mts.	CA	USA	36.2	-117.1	6/7/2014	M
Junco h. thurberi x caniceps	5	Pamint Mts.	CA	USA	36.2	-117.1	6/7/2014	M
Junco h. thurberi x caniceps	5	Pamint Mts.	CA	USA	36.2	-117.1	6/8/2014	M
Junco h. thurberi x caniceps	5	Pamint Mts.	CA	USA	36.2	-117.1	6/8/2014	M
Junco h. thurberi x caniceps	5	Pine Creek Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Pine Creek Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Pine Creek Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Pine Creek Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Pine Creek Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA				
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA	36.3	-115.7	4/26/2012	M
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA	36.3	-115.7	4/26/2012	M
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA	36.3	-115.7	4/27/2012	M
Junco h. thurberi x caniceps	5	Spring Mts.	NV	USA	36.3	-115.7	4/27/2012	M
Junco h. thurberi x caniceps	5	White Mts.	CA	USA	37.5	-118.2	6/6/2014	M
Junco h. thurberi x caniceps	5	White Mts.	CA	USA	37.5	-118.2	6/6/2014	M
Junco h. thurberi x caniceps	5	White Mts.	CA	USA				
Junco h. thurberi x caniceps	5	White Mts.	CA	USA				
Junco h. thurberi x caniceps	5	White Mts.	CA	USA				
Junco h. thurberi x caniceps	5	White Mts.	CA	USA				
Junco h. thurberi x caniceps	5	White Mts.	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA	37.4	-118.2	5/3/2012	M
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA	37.4	-118.2	5/3/2012	M
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA	37.4	-118.2	5/3/2012	M
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				

Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco h. thurberi x caniceps	5	Wyman Canyon	CA	USA				
Junco hyemalis caniceps	5	Toiyabe	NV	USA	39.3	-117.1	4/28/2012	M
Junco hyemalis caniceps	5	Toiyabe	NV	USA	39.3	-117.1	4/28/2012	M
Junco hyemalis caniceps	5	Toiyabe	NV	USA	39.3	-117.1	4/29/2012	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.4	6/29/2014	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis thurberi	5	Tahoe	CA	USA	38.8	-120.3	6/29/2014	M
Junco hyemalis oreganus	5	Susan Island	BC	Canada			8/31/2004	F
Junco hyemalis oreganus	5	N. Banks Island	BC	Canada			9/2/2004	M
Junco hyemalis oreganus	5	Porcher Island	BC	Canada			9/3/2004	M
Junco hyemalis pontilis	5	Sierra Juarez	BC N	Mexico			5/23/2013	M
Junco hyemalis oreganus	5	Porcher Island	BC	Canada			9/3/2004	M
Junco hyemalis pontilis	5	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/23/2013	M
Junco hyemalis pontilis	5	Sierra Juarez	BC N	Mexico	32.0	-115.9	5/25/2013	M
Junco hyemalis pontilis	5	Sierra Juarez	BC N	Mexico			5/25/2013	M
Junco hyemalis townsendi	5	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/27/2013	M
Junco hyemalis townsendi	5	San Pedro Mártir	BC N	Mexico	31.0	-115.5	5/28/2013	M
Junco hyemalis townsendi	5	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis townsendi	5	San Pedro Mártir	BC N	Mexico	31.0	-115.6	5/28/2013	M
Junco hyemalis montanus	5	Wallowa NF	OR	USA	45.3	-117.1	6/17/2013	M
Junco hyemalis montanus	5	Wallowa NF	OR	USA	45.3	-117.1	6/17/2013	M
Junco hyemalis montanus	5	Wallowa NF	OR	USA	45.3	-117.1	6/17/2013	M
Junco hyemalis pinosus	5	Nascimiento Road	CA	USA	36.0	-121.5	7/5/2014	M
Junco hyemalis pinosus	5	Santa Cruz Mts.	CA	USA			7/3/2014	M
Junco hyemalis pinosus	5	Nascimiento Road	CA	USA	36.0	-121.4	7/5/2014	M
Junco hyemalis pinosus	5	Nascimiento Road	CA	USA	36.0	-121.4	7/6/2014	M
Junco hyemalis caniceps	5	Aspen	UT	USA	39.3	-106.6	6/16/2014	M
J. h. caniceps x dorsalis	5	Mount Taylor	NM	USA	35.3	-107.6	6/21/2014	F
Junco hyemalis dorsalis	5	Sacramento Mts.	NM	USA			6/25/2014	M
Junco hyemalis dorsalis	5	Sacramento Mts.	NM	USA			6/25/2014	M
Junco hyemalis dorsalis	5	Sacramento Mts.	NM	USA	32.9	-105.8	6/26/2014	M
Junco hyemalis dorsalis	5	Dude Mountain	AZ	USA				
Junco phaeonotus palliatus	5	Sierra Madre Occ.	CHIH	Mexico			6/11/2004	M

The great diversity of life forms in our planet has challenged our understanding of the natural world and indeed the origin of our own species for centuries. Charles Darwin's decisive discovery of evolution through natural selection and the later development of the modern evolutionary synthesis stablished the theoretical basis of the mechanisms underlying the existence of discontinuous biological units. As a result, they provided an unequivocal explanation of the origin of biodiversity and of our relationship with the reality of the physical universe, shaping the modern thought like no other scientific theory in history, in words of Jacques Monod. The scientific field of Evolutionary Biology, and the resulting, ever-since thriving corpus of knowledge, is but the will of men and women in achieving a better comprehension of the mystery. Aiming to contribute to this knowledge, the objective of this dissertation is to reconstruct the evolutionary history and to study the evolutionary mechanisms involved in the diversification of the genus *Junco* (Aves: Emberizidae), a case of rapid phenotypic differentiation across North America that provides a unique opportunity to test the relative roles of neutral and selective mechanisms in driving lineage divergence and speciation.

mncn museonacionaldecienciasnaturales
CSIC Consejo Superior de Investigaciones Científicas
UAM Universidad Autonoma de Madrid